THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

The Role of Sugars for Protein Stabilization

CHRISTOFFER OLSSON

CHALMERS

Department of Physics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2018

The Role of Sugars for Protein Stabilization
CHRISTOFFER OLSSON

ISBN 978-91-7597-808-6

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 4489
ISSN 0346-718X

Department of Physics
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
Telephone + 46 (0)31-772 1000

Cover illustration: Preferentially hydrated myoglobin. Artistic interpretation by Erik Widell.

Chalmers Reproservice
Göteborg, Sweden 2018

The Role of Sugars for Protein Stabilization

Christoffer Olsson
Department of Physics
Chalmers University of Technology

Abstract

The understanding of biomolecular interactions with water and co-solutes can lead to greater knowledge regarding the mechanisms behind biomolecular stabilization. This is highly important for developing technologies aimed to preserve biological materials. Such techniques include cryopreservation of pharmaceuticals or human organ transplants, for example. For these purposes, the disaccharide trehalose has been shown to be an outstanding biomolecular stabilizing agent during cryostorage or storage of desiccated materials.

In this thesis, the questions regarding the stabilizing role of trehalose is addressed from several different angles. Structural properties of trehalose in water are studied and are compared to those of a similar sugar molecule, namely sucrose. From these studies it was concluded that there were surprisingly small differences between the interactions of trehalose or sucrose with water. The thermodynamic properties of trehalose–water–protein systems were investigated using DSC, where it was indirectly found that the protein hydration shell was not substituted by trehalose, and that the protein stability did not necessarily couple to the glass transition temperature of the trehalose–protein–water-matrix. The structure and dynamics of such a ternary trehalose–water–protein system was also investigated using neutron diffraction combined with EPSR, and QENS combined with an MD simulation. In these studies, it was primarily found that the trehalose molecules were preferentially excluded from the protein surface, and that the local motions of the protein residues were slowed down via a reduction in the motion of the water molecules at the protein surface. Furthermore, the temperature dependences of relaxation dynamics in this system were measured using dielectric spectroscopy. This study showed that the presence of protein hinder certain local trehalose motions, and that the relatively slow dynamics of the trehalose solvent governs the conformational motions of the protein.

The presented results elucidates some fundamental properties of how proteins and trehalose behave and interact, which may benefit the development of new biomolecular protective co-solutes.

Keywords: *trehalose, protein, biomolecule, water, amorphous, cryopreservation, neutron scattering, neutron diffraction, EPSR, DSC*

List of publications

This thesis is based on the work contained in the following papers:

Paper I. Structure of aqueous trehalose solution by neutron diffraction and structural modeling
Christoffer Olsson, Helén Jansson, Tristan Young, and Jan Swenson
J. Phys. Chem. B ***2016****, 120, 12669-12678*

Paper II. Structural comparison between sucrose and trehalose in aqueous solution
Christoffer Olsson and Jan Swenson
In Manuscript

Paper III. The Role of Trehalose for the Stabilization of Proteins
Christoffer Olsson, Helén Jansson, and Jan Swenson
J. Phys. Chem. B ***2016****, 120, 4723-4731*

Paper IV. Structural Role of Trehalose for Protein Stabilization and Inhibiting Protein Aggregation from Neutron Diffraction
Christoffer Olsson and Jan Swenson
Submitted to Physical Chemistry Chemical Physics

Paper V. Mechanism of Trehalose Induced Protein Stabilization from Quasielastic Neutron Scattering and Molecular Dynamics Simulations
Christoffer Olsson, Victoria Garcia-Sakai, Samuel Genheden, and Jan Swenson
Submitted to Physical Chemistry Chemical Physics

Paper VI. Dielectric spectroscopy study of proteins embedded in trehalose and water
Christoffer Olsson, Rano Zangana, and Jan Swenson
In Manuscript

Publications not included in the thesis

Paper VII. Motions of water and solutes - slaving versus plasticization phenomena

Izaskun Combarro Palacios, Christoffer Olsson, Christina S. Kamma-Lorger, Jan Swenson, and Silvina Cerveny.

Submitted

My contributions to the papers:

Paper I

I prepared all the samples and performed all the diffraction experiments together with the co-authors. I performed the data processing, EPSR simulations, data analysis, and I was the main author of the manuscript.

Paper II

I prepared all the samples and performed all the diffraction experiments together with the co-author. I performed the data processing, EPSR simulations, data analysis, and I was the main author of the manuscript.

Paper III

I performed sample preparations and DSC measurements. Furthermore, I did most of the data analysis, and I wrote the first draft of the manuscript.

Paper IV

I prepared all the samples and performed all the diffraction experiments together with the co-author. I performed the data processing, EPSR simulations, data analysis, and I was the main author of the manuscript.

Paper V

I prepared all the samples and performed all the diffraction experiments together with the J.S. and V.G.S. I performed the experimental data processing, and data analysis. S.G. set up and ran the MD simulations, and I did the analysis of these simulations. I was also the main author of the manuscript.

Paper VI

I performed most of the sample preparations and BDS measurements, with help in part from R.Z.. Furthermore, I did most of the data analysis, and I was the main author of the manuscript.

Glossary

BDS	**B**roadband **D**ielectric **S**pectroscopy
DSC	**D**ifferential **S**canning **C**alorimetry
EP	**E**mpirical **P**otential
EPSR	**E**mpirical **P**otential **S**tructure **R**efinement
HB	**H**ydrogen **B**ond
MD	**M**olecular **D**ynamic
MMC	**M**etropolis **M**onte **C**arlo
NIMROD	**N**ear to **I**nter **M**ediate **R**ange **O**rder **D**iffractometer
QENS	**Q**uasi **E**lastic **N**eutron **S**cattering
RMC	**R**everse **M**onte **C**arlo
RP	**R**eference **P**otential
$\mathbf{T}_g$	Glass transition temperature
$\mathbf{T}_d$	Denaturation temperature

Contents

1

Introduction

In the end of May 2013, pictures of a frozen wholly mammoth began to pop up in the newspapers. A team of Russian and South Korean scientists had discovered this extinct animal in a sheet of ice in Siberia; an occurrence which has become relatively more common recently. What was so spectacular about this particular mammoth however, was that it was possible to extract blood and seemingly fresh meat from it, after it had been frozen for over 43 000 years.[1] This fascinating discovery could very well mean that it could be possible to extract DNA from an intact mammoth cell, which could furthermore lead to the cloning of an animal that went extinct around 4000 years ago. Shortly after this discovery, I began my PhD-studies aimed to study the fundamentals of cryopreservation. Thus, this story about the mammoth served as a great source of inspiration; what made this particular mammoth-carcass able to be so well preserved for such a long time? How come biological material in general degenerate over time? What are the mechanisms responsible for suppressing these degenerations? There are multitudes of answers to these questions, most of them reaching far beyond the scope of any single thesis, but at least some of these questions will be addressed here.

The process of preserving biological material has huge importance in a large number of different fields. For example, the field of preserving biological compo-

nents for human transplants, such as body tissues, or blood, has the capabilities of saving many lives. Improved preservation techniques may have a special increase of demand in the future due to advances in e.g. tissue engineering,[2] where appropriate storage of the created tissues is needed. In other areas such as pharmaceutics or food-production, it is of course also important to obtain a long shelf-life for all the different products. Similarly, in several areas within biotechnology it is vital to keep cell cultures or other biological material viable for longer periods of time.

Two of the most common methods of preserving biological material are cryopreservation and freeze-drying.[3] Cryopreservation is the term used to describe preservation of materials at low temperatures. This technique is advantageous due to the decrease of molecular motions of the stored complex biomolecular structures; motions which normally are required for biological function, but are also capable of disrupting or breaking these structures. However, a huge disadvantage of this temperature decrease is the formation of ice, which often cause great damage to the preserved material.[4] Freeze-drying on the other hand attempts to immobilize the biological material by removing the water around it. Without an aqueous medium, most biological processes cease, and the biomolecules become stabilized. This process can however also damage the biological molecules and is far from optimal in many scenarios.[5,6] The problems connected to cryopreservation or freeze-drying are commonly, at least in part, dealt with by the addition of protective molecules (cryoprotectants and lyoprotectants). There are many such molecules, each typically well suited for one of the two methods, but there is one molecule that excels in stabilizing biological material when it comes to both mentioned methods: trehalose.

Trehalose is a sugar-molecule, very similar to more common sugar-molecules such as sucrose. Although sucrose and other sugar molecules typically exhibit good stabilization properties,[7] trehalose almost always has the superior properties (see e.g. Ref. 8). Why trehalose possess these superior properties is still

unclear, and it is the aim of this thesis to provide insights into this question.

For this purpose, the work in this thesis investigates different aspects about trehalose. In paper I, the structural properties of trehalose in an aqueous solution (33 wt% trehalose) is studied, and in paper II this work is compared to a similar study of sucrose with the same water concentration. The work in these two papers was done by neutron diffraction and empirical potential structure refinement-modeling. Papers III–VI concerns the interaction of trehalose with water and protein. Thermodynamic properties, such as glass transition- and denaturation temperatures, was studied by the use of differential scanning calorimetry in paper III. The molecular structure of this three-component system was studied by the use of neutron diffraction and empirical potential structure refinement-modeling in paper IV, and the dynamics of the same system was investigated using quasielastic neutron scattering and molecular dynamic simulations in paper V. The dynamics of a set of drier systems, and thus slower dynamics, was studied in paper VI, by the use of broadband dielectric spectroscopy.

In the structural diffraction study of water with trehalose (paper I), the interaction between water and trehalose was studied. It was shown that plenty, although relatively weak, hydrogen bonds between trehalose and water forms, and also that trehalose has a very small probability of forming intermolecular clusters. As a comparison to the results for trehalose obtained in paper I, sucrose was also studied, using the same methods, in paper II. There it was shown that trehalose and sucrose exhibit quite similar structure when dissolved to this particular concentration. A similar amount of hydrogen bonding, and a similar lack of clustering is shown, however there are indications of that trehalose exhibits a larger destructuring effect on the bulk-like structure of water. In the differential scanning calorimetry study (paper III), evidence for the pref-

erential hydration model was found. The study also includes glass transition and denaturation temperatures for a wide concentration range of both trehalose and protein. Evidence for the preferential hydration model was also found in paper IV and paper V for a more hydrated system than those generally studied in paper III. Paper IV also shows how trehalose inhibits protein–protein interactions, and how water interacts with trehalose in the presence of proteins. In paper V it is also shown how the presence of trehalose slows down certain protein motions.

This thesis contains background information about the materials and the experimental techniques that are used in these studies. Chapter 2 gives an overview of liquids and amorphous materials and how these can be investigated. This is followed by a more detailed description of the specific molecules used for these studies: water, myoglobin, sucrose, and trehalose. Subsequently, some models regarding interactions between these materials, and how protein stabilization occurs, are presented.
Chapter 3 describes the theories behind the different techniques, and in chapter 4 it is explained how these techniques were applied in the presented studies. Chapter 5 briefly describes the obtained results from the papers, and chapter 6 gives a summary of this work and an outlook on how to proceed with answering the questions concerning the stabilizing role of trehalose.

2

Materials

2.1 Liquids and amorphous materials

When discussing the structure of a liquid or an amorphous material, a specific framework is necessary to define what is meant by structure. As opposed to the structure of a crystal, where all the atoms and molecules have a more or less well-defined position relative to one another, an amorphous material lack most of such structural ordering. In fact, the definition of an amorphous material is a material which lack long-range order between atoms, i.e. there are no repeating units that can be found throughout the material.[9] Of course, the lack of long-range order does not mean that there is no structure at all. On shorter length scales, molecules in an amorphous material can exhibit very complex and interesting structures which determine the properties of the material, such as clusters, vesicles and ring structures, to name a few examples. The materials studied in this thesis appear in two types of amorphous phases, liquids and glasses. This section gives a brief general overview of these two phases concerning their general properties and how their structures can be described.

2.1.1 From liquids to glasses

Typically, when a liquid is cooled the viscosity increases, partly due to the decrease in kinetic motion of the individual particles in the liquid, and partly due to an increase in density. When the motion becomes slower the particles propagate through the material at a slower rate – they become "caged" by the surrounding molecules. Thus when a force is applied to the liquid it takes longer time for the particles to react to that force. This phenomenon is what we typically experience as a high viscosity. When the liquid is cooled below its freezing point it becomes energetically favorable for the particles to form lattices of particles, a process known as crystallization. However, for crystallization to occur the particles need some time to reorient themselves into the correct position of a lattice site. Thus if the freezing occurs very fast (quenching) the viscosity increases rapidly – the movements of the particles quickly slow down – and the time required for the particles to arrange themselves into a crystal lattice-structure may grow to extremely long times. When these time spans become larger than the experimental time spans (i.e. such that particle motions are too slow to react during the time span of the experiment) the material can go through what is called the *glass transition*; the material becomes vitrified.

The glass transition is often detected by e.g. calorimetric measurements, and can be seen as a gradual change of the enthalpy (H) dependent on temperature (Figure 2.1). This change in the slope of the enthalpy shows that the glass transition behaves rather differently compared to a "normal" transition, such as a crystallization process, where the enthalpy changes abruptly at a certain temperature (melting/crystallization-temperature T_m/T_c).*
A transition that exhibit a discontinuity in the directly observable thermodynamic quantities, such as enthalpy (and also volume and entropy), is called a first-order transition. A glass transition however displays a change in the slope of these thermodynamic properties (Figure 2.1), which means that the

*Other thermodynamic properties, such as the thermal expansion behaves similar to the heat capacity.

derivatives of these thermodynamic quantities – such as the thermal expansion $\alpha_T = \frac{\delta lnV}{\delta T}$, or the heat capacity $C = \frac{\delta H}{\delta T}$ – are discontinuous at the glass transition. This type of transition is a so called second-order transition.[9,10]
The temperature at which these discontinuities occur often serves as the definition of the glass transition temperature (T_g). An important difference between this transition temperature and that of for example a crystallization temperature, is that the glass transition temperature depends on the history of the material. For example, T_g will typically become lower if the material is cooled at a slower rate. This is due to the extra time available for the material to reach equilibrium at a slower cooling rate. If the material is cooled at a greater rate, the particles in the material become immobilised before they are able to reach a new equilibrium.[9]

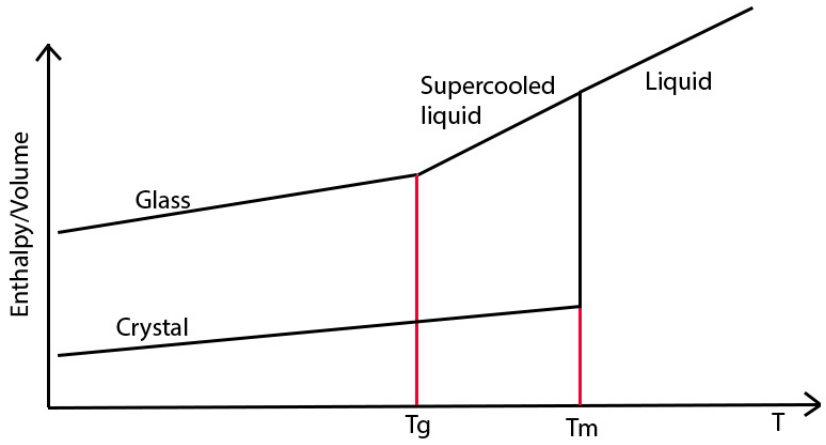


Figure 2.1: Graph showing how a substance varies in volume or enthalpy depending on temperature during transitions from a liquid state into either a glassy state or into a crystal.

A common definition for the glass transition is that it occurs when a material reaches a viscosity of 10^{13} poise. At this viscosity the material can be regarded as a solid for all practical purposes. Another common directly related definition for the glass transition is that it occurs at the temperature where the characteristic α-relaxation process of the material, which is related to co-

operative motions of the particles, reaches a relaxation time of 100 seconds. As previously mentioned, when a particle in a liquid is perturbed by a force it can only move if there are available sites for the particle to move to. The α-relaxation time (τ_α) is thus the characteristic time it takes for several particles to perform such cooperative motions. In a low viscosity liquid, τ_α is short and thus the particles collectively move in response to an applied force, but as τ_α increases the longer it takes for these collective motions to occur, and thus the material appears as more and more viscous. τ_α can typically be probed using techniques such as dielectric spectroscopy.

2.1.2 Structure of Liquids and Glasses

In order to discuss the structure of a liquid there is a need for establishing a formalism that can accurately describe this. The most common and simple way to do this is via the pair correlation function, $g(r)$*. This function is essentially the probability of finding another atom within a spherical shell between r and $r + dr$ at a distance r from the center of any arbitrary atom, as illustrated in Figure 2.2.

In a crystalline material this function show sharp peaks at specific r values, corresponding to the well-defined atomic distances within a crystal (bottom figure in Figure 2.2). The broadening of these peaks is mainly due to structural thermal fluctuations of the atoms. The pair correlation function for an amorphous material (upper figure in Figure 2.2) on the other hand oscillates smoothly around unity at relatively large distances from the center. This reflects the random orientation of atoms at large distances from any given atom, where it is expected to find a number of atoms within a given volume, $n(r)$, equal to that of the average number density (ρ) of the material ($n(r) = \rho g(r)$). The peaks in an amorphous material at short distances are generally also broader compared to those in a crystalline structure. This is mainly due to

*More specifically, the *static* pair correlation function.

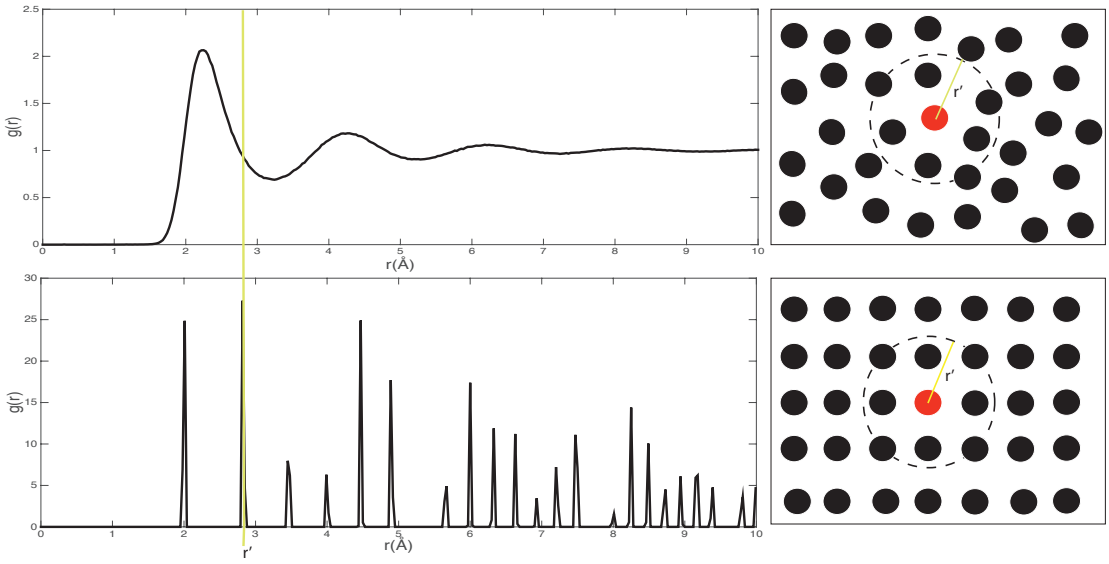


Figure 2.2: Pair correlation functions. Upper figure shows the pair correlation function of a hard-sphere liquid. Lower figure shows the pair correlation function of a crystal structure.

that the bonding sites are less well-defined in an amorphous material, but also partly due to a higher degree of structural thermal fluctuations of the atoms.[9] The pair correlation function originates from the autocorrelation function of the local atomic density $n(\mathbf{r})$, defined as[11]:

$$n(\mathbf{r}) = \sum_i \delta(\mathbf{r} - \mathbf{r}_i) \tag{2.1}$$

Where δ is the Dirac delta function, and $\mathbf{r}_i$ is the position of atom i. The autocorrelation of $n(\mathbf{r})$ is*:

$$G(\mathbf{r}) = \frac{1}{N} \int n(\mathbf{r}')n(\mathbf{r}' + \mathbf{r}))d\mathbf{r} = \frac{1}{N} \sum_{ij}^{N} \delta(\mathbf{r} + \mathbf{r}_j - \mathbf{r}_i) \tag{2.2}$$

and by separating the summation in two terms, one where $i = j$ and one where

*This autocorrelation function can be interpreted as a measure of the degree of correlation between two atoms separated by a distance $\mathbf{r}$.

$i \neq j$, this can be reduced to:

$$G(\mathbf{r}) = \delta(\mathbf{r}) + \frac{1}{N}\sum_{i\neq j}^{N}\delta(\mathbf{r}+\mathbf{r}_j-\mathbf{r}_i) = \delta(\mathbf{r}) + \rho g(\mathbf{r}) \tag{2.3}$$

where the relationship $\rho g(\mathbf{r}) = \frac{1}{N}\sum_{i\neq j}\delta(\mathbf{r}+\mathbf{r}_j-\mathbf{r}_i)$ is the definition of $g(\mathbf{r})$ and refers to the pair correlation function (i.e. the correlation between any two different atoms), whereas the first Dirac delta term in equation 2.3 is the self-correlation part, i.e. how an atom correlates with itself in space.

Multicomponent systems

When a material contains more than one atom type it is often useful to reduce the total pair correlation function into a weighted sum of so called partial pair correlation functions of different atom pairs. These are often denoted as $g_{\alpha\beta}(\mathbf{r})$, where α and β represent two different atom types, and is interpreted as a probability of finding an atom of type β at a distance r from an atom of type α (or vice versa since $g_{\alpha\beta}(\mathbf{r}) \equiv g_{\beta,\alpha}(-\mathbf{r})$). The sum, describing the autocorrelation function is then:[11]

$$G(\mathbf{r}) = \delta(\mathbf{r}) + \rho g(\mathbf{r}) = \sum_{\alpha} c_\alpha \delta(\mathbf{r}) + \rho \sum_{\alpha,\beta\geq\alpha}(2-\delta_{\alpha\beta})c_\alpha c_\beta g_{\alpha\beta}(\mathbf{r}) \tag{2.4}$$

Where c_α and c_β are the fractions of atoms α and β in the sample respectively. The partial pair correlation functions are very useful for the analysis of an amorphous material since the sum of them completely describes the average structure of the material as one single function, and the individual pair correlation functions captures specific average structures between different molecular species. Various information which can be extracted will be discussed in more detail in section 4.2. Section 3.1.1 describes how to obtain the pair correlation functions by neutron scattering and modeling techniques.

Dynamic correlation function

The correlation function discussed so far assumed stationary particles, however this formulation can be expanded to include displacements over time. The obtained dynamic pair correlation function, called the van Hove function, is often written as*:

$$G(\mathbf{r},t) = \frac{1}{N}\sum_{i,j}^{N}\delta(\mathbf{r}+\mathbf{r}_i(0)-\mathbf{r}_j(t))d\mathbf{r} \tag{2.5}$$

This should be viewed as the probability density of finding a particle at position $\mathbf{r}$ at time t, given that this or another particle was at the origin at time 0. Again, separating this function into its pair-correlation function ($G(\mathbf{r},t)_{Distinct}$) and its self-correlation function ($G(\mathbf{r},t)_{Self}$), the following two equations can be written:

$$G(\mathbf{r},t)_{Distinct} = \frac{1}{N}\sum_{i,j\neq i}^{N}\delta(\mathbf{r}+\mathbf{r}_j(0)-\mathbf{r}_i(t)) \tag{2.6}$$

$$G(\mathbf{r},t)_{Self} = \frac{1}{N}\sum_{i}^{N}\delta(\mathbf{r}+\mathbf{r}_i(0)-\mathbf{r}_i(t)) \tag{2.7}$$

It is easy to see that the different van Hove functions reduce to the static case when $t = 0$.

2.2 Water

Equipped with the tools to characterize the structure of a liquid, the formalism described above will be exemplified in this section using a relatively simple compound, one of the most abundant substances on earth: water. This is – to no surprise – a well-studied material, due to its presence in a vast number of chemical and biological reactions. For the scope of this thesis, it is important to highlight some of the properties of water, since it is the medium in which all

*Or rather, the real part of the van Hove function[12]

the studied materials were dissolved in. Furthermore, the structure of water is also highly relevant for a part of the study presented in papers I and II.

2.2.1 Properties

Water consists of one oxygen atom covalently bonded to two hydrogen atoms. The hydrogen atoms are bonded to the oxygen in a triangular shape, where the H–O–H angle is around 104.5° on average.[13] This geometry stems from how the electron density in a water molecule is slightly higher at the oxygen atom and slightly lower at the hydrogen atoms, which also gives the molecule its dipole moment. Water molecules attract each other through hydrogen bonds (HBs) between hydrogens and oxygens due to this resultantly small charge difference. Compared to the covalent intramolecular bonds, the intermolecular hydrogen bonds are relatively weak. However, these hydrogen bonds are widely believed to be the main reason for a series of peculiar properties of water.

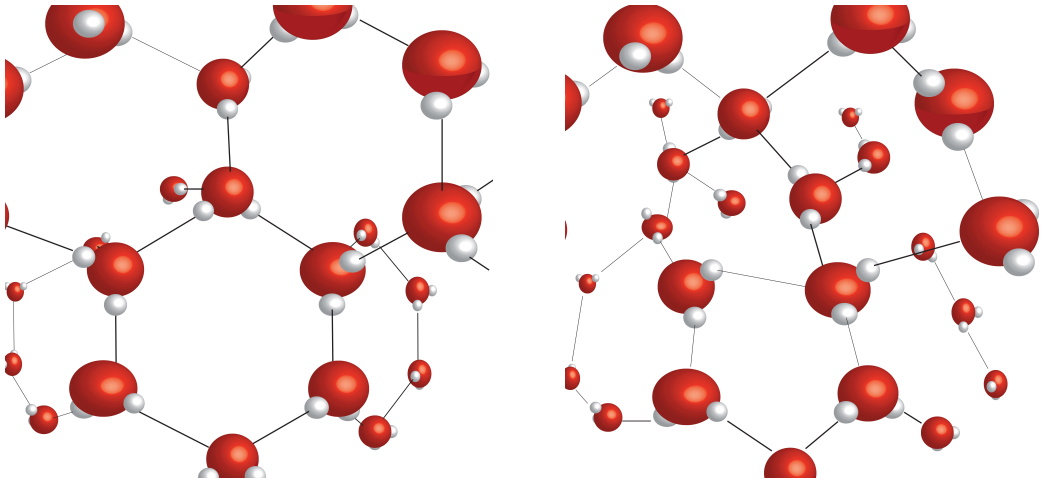
Water is in many aspects not a normal material, but possesses several anomalous properties. Most commonly known is probably the density maximum of water at 4°C; this property means that for a certain temperature interval, the density of H_2O *decreases* with decreasing temperature, as opposed to almost every other known substance.[13] Another anomalous property of water is its high specific heat capacity (4.18J $g^{-1}K^{-1}$ at 25°) which is one of the highest heat capacities out of all known substances.*

To understand these anomalous properties, the hydrogen bonded networks and structures formed within water at its different phases (ice and liquid) is discussed below.

*There are of course a lot more anomalous properties of water which are not mentioned here. For more information, the reader is referred to e.g. Ref. 13

2.2.2 Structure

The intermolecular structure of ice is typically shown as in Figure 2.3 a. Each water molecule has four other water molecules as first-order neighbors (as discussed in section 4.2.1),[13] forming a tetrahedral structure in three dimensions. Ice is in a crystalline (relatively*) well-structured state; each water molecule is in a well-defined lattice point, connected symmetrically to four other water molecules. Liquid water is not as symmetrical since the water molecules translate and rotate due to thermal fluctuations. One widely accepted view of water structure has been that water molecules form on *average* a tetrahedral network, where each water molecule, on average, binds to 4.4 other water molecules.[14] These bonds constantly break and form with a lifetime of ∼1 ps at room temperature.[15]



(a) Ice (b) Water

Figure 2.3: Schematic of ice and water structure.

The structure of regular bulk water can be studied by the partial pair correlation functions, O–O, O–H, and H–H (Figure 2.4). Particularly $g_{OO}(r)$ – the partial pair correlation function between two water-oxygen atoms – tell us something interesting regarding the coordination numbers of water at 298K compared to ice at 220K. The first sharp peak at 2.8 Å indicates the typical

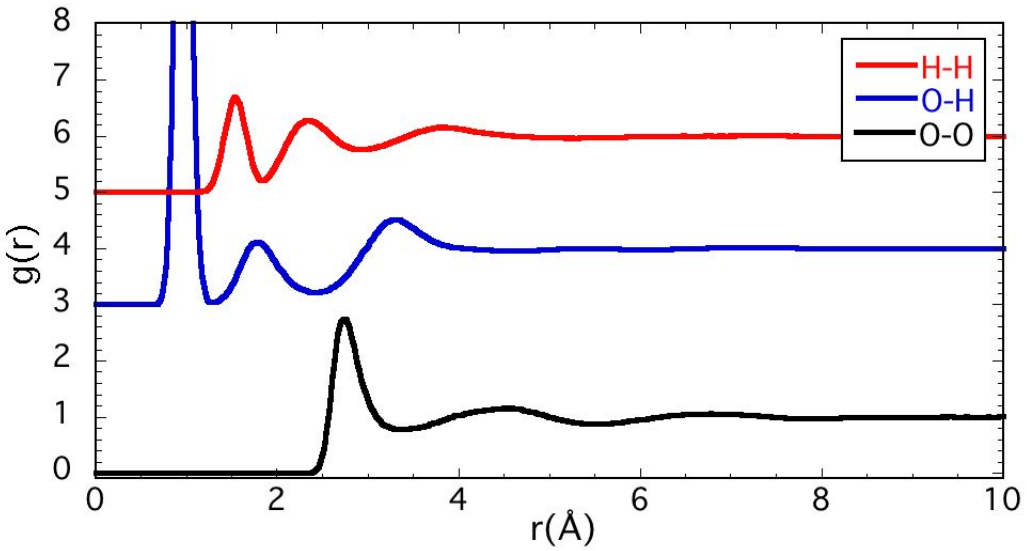
*Ice is actually not specifically structured compared to other crystalline materials.

distance to the first-order neighbor of a water molecule. Notably, this peak, which shows the first-order distance, moves only to slightly higher values and becomes somewhat broader as the ice melts. A much more pronounced effect however is seen for the second peak (i.e. the second-order coordination shell), at about 4.5 Å, which is highly pronounced in the ice structure and is significantly less pronounced in liquid water. This second peak is seen as a signature of tetrahedral structure in water and ice. In ice, the water molecules are pretty well coordinated at the second-order neighboring (and to third and fourth etc. as well, but to a lower and lower extent). In liquid water this coordination is still present, however less defined, particularly at higher temperatures.* [16,17]

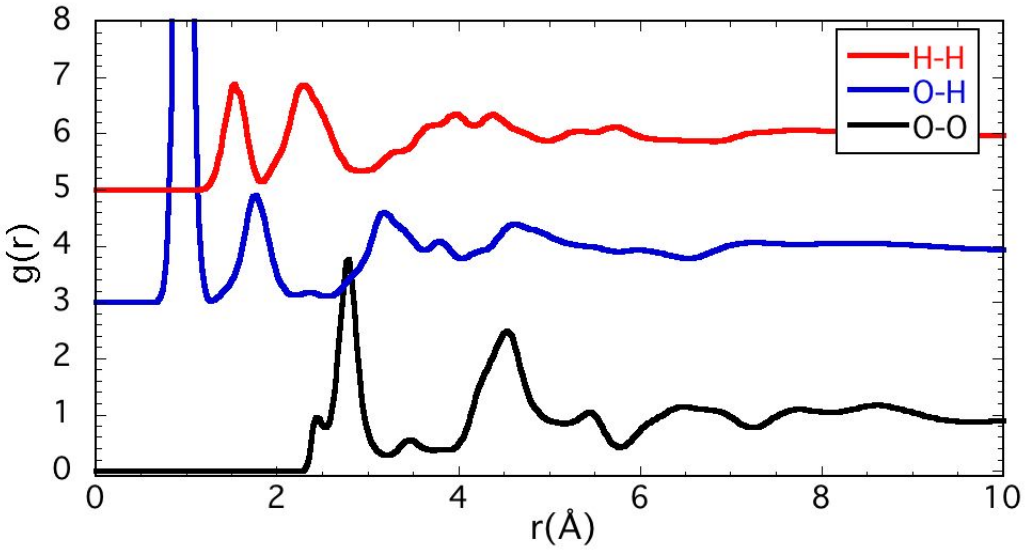
From these correlation functions and other similar derived quantities, it has been shown that in liquid water a water molecule is typically surrounded by a tetrahedral structure of neighboring water molecules. The molecules in these structures are not stationary but rather fluctuate in their positions. First order neighbors exchange with second-order neighbors, and interstitial water molecules break up the otherwise four-coordinated water into five-coordinated water. [18,19]

The abnormal density maximum of water can thus be explained by noting that the existence of destructured water arrangements increase with increasing temperature, and the destructured water is more densely packed than the structured ones, even though both types of structures expand in volume with increasing temperature. Hence, at a certain point (at $\sim$4°C) the sum of the existence of density-increasing (destructured) water structures and the density increasing effect of lowering the temperature reaches a maximum. Below this temperature, more tetrahedral structures begin to form, thus lowering the density more than the decreasing temperature increases it. [13]

*A similar trend is found for pressure dependence; higher pressure changes the partial pair correlation functions similar to higher temperatures. [16,17]



(a) Water at 298K



(b) Ice at 220K

Figure 2.4: Partial correlation functions for O–O, O–H, and H–H correlations of (a) water at 298K and (b) ice at 220K. Data obtained from the ISIS disordered materials database.[17,20]

2.3 Disaccharides

Disaccharides are very common molecules in biological materials, and are used in many different biochemical reactions. The main focus of investigation in this thesis is to improve stabilization of biological molecules, and for this purpose

the disaccharide trehalose has shown many promising effects. Trehalose appear to be present in a wide range of different extremophiles, and it has been shown in countless studies[8,21,21–30] that trehalose possess an extraordinary ability to stabilize biological material against many different types of environmental stresses, such as desiccation, extreme temperatures or extreme pressure (see for example Ref. 31 wherein the extremophile tardigrade is studied, a small microscopic animal capable of surviving some extreme environments). Thus, in order to understand the general mechanism behind biopreservation, the specific interactions between trehalose and proteins have been investigated. However, trehalose is by no means unique in its ability to stabilize proteins. For example, the more common disaccharide, sucrose, has a very similar chemical structure to trehalose and also exhibit many stabilizing properties, although not to the same extent as trehalose. By comparing the differences between properties of these two disaccharides one may thus obtain a more detailed answer as to what mechanisms yield the most biomolecular stabilization. Before going into the details regarding their role in biomolecular stabilization, this section will however focus more on general properties of these two disaccharides.

2.3.1 Chemical Properties

Trehalose and sucrose have identical chemical formulas ($C_{12}H_{22}O_{11}$), with a molecular weight of 342.296Da. They both consist of two monosaccharides connected by a glycosidic linkage; trehalose is built up by two glucose rings, and sucrose by one glucose and one fructose rings. A simple cartoon of their molecular structure can be seen in Figure 2.5. Their chemical formula are furthermore identical to a number of other disaccharides, such as lactose and maltose. In fact, many of the disaccharides, are very similar in their structures, apart from the positioning of the different hydroxyl and hydroxy-methyl groups on the monosaccharide rings. The similar disaccharides share many properties with trehalose, however the small structural differences have been shown to

play a highly important role for the molecular functions studied in the work presented in this thesis.

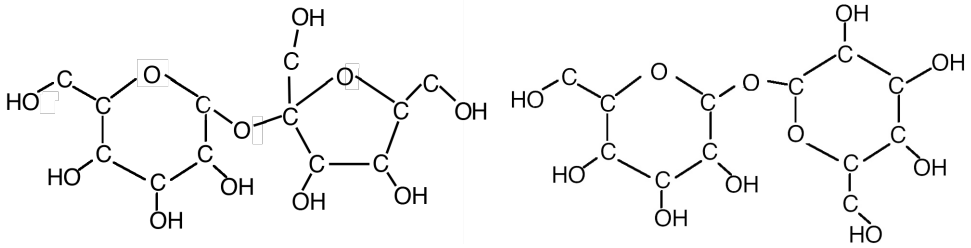


Figure 2.5: Molecular structure of sucrose (left) and trehalose (right)

The melting point of water-free (anhydrous) trehalose is 203∘C, whereas sucrose exhibits a more complex behavior as it decomposes at 186°C into its two monosaccharides, which are both in liquid form at that temperature*. Trehalose easily binds to two water molecules – forming dihydrate crystals – which is the most stable form of trehalose at ambient conditions. This form has a melting temperature of 97°C and, after being melted, dissociates its water molecules and forms the anhydrous solid phase at about 130°C.[32]

Comparing with similar disaccharides, trehalose has the highest T_g.[33] This property is highly dependent on water concentration, as the water molecules are able to plasticize the sugar matrix. It is also widely believed that trehalose not only forms a glass at the highest temperature, but also that it is superior in remaining in a stable glassy state.[34,35] This effect stem from the fact that trehalose can transform into the dihydrate phase when exposed to water. A small water addition to any other glassy disaccharide matrix destabilizes the entire glass by homogeneously distributing the water molecules. This process leads to a decrease in T_g which subsequently leads to a lower stabilization effect (see e.g. Refs. 26 and 36). In a glassy trehalose matrix however, the molecules which are immediately exposed to water acts as a buffer by absorbing it and thus excluding the water from the remaining glass-matrix. This process

*This process is even more complex, including a caramelization-process which will not be discussed.

makes T_g more stable and therefore provides a more reliable stabilization.[35] Other chemical properties of interest between these two disaccharides worth mentioning include: Sucrose is much more soluble than trehalose; 1 g of water dissolves as much as 2.14 g sucrose, but only 1.1 g trehalose.[37] Sucrose is twice as sweet-tasting as trehalose, which is a property that has been suggested to be related to its higher solubility[38].

2.3.2 Structure

Several studies have investigated intra- and inter-molecular clustering of disaccharides in solution.[39–42] The literature does however not agree on all issues, and thus the current section only gives a brief overview of a selection of results.

Intramolecular structure

The structure of trehalose in an aqueous solution is highly dependent on the water concentration. As the amount of water decreases, trehalose tend to fold across its glycosidic linkage, forming intramolecular bonds between the two glucose rings.[40] This folding prevents intermolecular binding between the water and trehalose due to that many of the hydroxyl groups on the trehalose are occupied in the intramolecular fold. With an increasing amount of water, the folded structure breaks up and subsequently makes more hydroxyl groups available for further interactions between water and trehalose.[40] For more diluted cases it has been shown by molecular dynamics simulations that trehalose exhibits a larger flexibility than sucrose, which is likely a consequence of trehalose's reluctancy to form intramolecular hydrogen-bonds compared to sucrose.[39]

Intermolecular structure

Intermolecular disaccharide interactions are other important aspects for the different properties of the disaccharides. In the dehydrated state (< 0.5 water

molecules per trehalose), trehalose has been shown[41] to exhibit a more homogeneous amorphous network than sucrose which form sucrose-rich crystalline domains from which the water is expelled. This property could be partly used to explain the superior stabilizing mechanism of trehalose during desiccation, since this finding suggests that trehalose is better at hosting larger molecules (such as proteins), and anchor them to the homogeneous glassy matrix via the residual protein water molecules.[41]

At high hydration levels (>38 water molecules per trehalose), trehalose also shows a high degree of homogeneity compared to sucrose, in the sense that the trehalose molecules do not cluster substantially.[39] This aspect is also found in paper I, where it is shown by neutron diffraction experiments that the clustering effect of trehalose is much smaller than previously reported.[39,43] In fact, according to the model in paper I, the trehalose molecules appeared to repel each other to a small extent. This was however also shown in paper II to be the case for sucrose at the same water concentration.

2.3.3 Effect on water structure

It was shown through a series of articles by Branca and Magazu et al[44–50]that the structure and dynamics of water is highly altered by the presence of trehalose. This effect has however been under some discussion,[51] which partly motivated the work done in paper I. In paper I it was also shown that trehalose perturb the bulk behavior of water by breaking up the tetrahedral network. By extensive – although not necessarily strong – hydrogen bonding, the water molecules are forced to reorient toward the trehalose hydroxyl groups, thus allowing for more interstitial water molecules in the first coordination shell of water. This also leads to a destruction of the second coordination shell (see Figure 3,4, and 5 in paper I). A revised model of trehalose was presented in paper II, which indicated not as strong effect on the water structure as in paper I. However, in paper II the destructuring effect on the water was compared

to that of water in sucrose, and it was found that the water structure indeed deviated more in the presence of trehalose than in the presence of sucrose, in support of e.g. Refs. 45, 48–50

The destruction of the water structure could be the main reason for why trehalose is good at depressing ice formation and why aqueous solutions of trehalose exhibits such a high T_g compared to similar disaccharides. However, other explanation models have been proposed. For example, it was shown that when a high concentration of the sugar-alcohol sorbitol (>70wt%) repress water crystallization, the water structure remained tetrahedral and, more similar to that of water in confined geometries.[52] It was therefore suggested that a network sorbitol molecules segregated water molecules into small water pockets, thus effectively creating confined water clusters. Water in such confinements are indeed also prevented from crystallizing, and exhibit slower dynamics[53,54]. Similarly, this effect has also been suggested to be the cause of why trehalose repress ice formation and reduce water mobility.[55]

2.4 Biomolecules

The term biomolecules typically refers to the molecules involved in biological systems and processes. They can be very complex, such as proteins or DNA, or much simpler, such as simple sugars involved in e.g. metabolism. When discussing biological stability, it is typically the more complex molecules and structures that are of interest since these are the ones more prone to breaking from environmental stress. The focus of this thesis is on the stabilization of proteins, but for a broader discussion it is important to sometimes deviate into the role of stability of other complex structures such as cell membranes.

2.4.1 Proteins

Proteins are the molecules that drives most biological processes. They are essential for the basic functions behind the processes of life, thus the loss of these functions are catastrophic for the organism hosting the proteins. Proteins are made up from chains of different amino acids, which link together according to different sequences in the DNA chain (i.e. genes). These protein chains typically fold to a subset of configurations that give the proteins their particular geometrical structures and functions. The folds and creases on the protein surfaces exhibit different chemical properties – such as being hydrophilic or hydrophobic – which typically determine their functions.[56]

Figure 2.6: 3D model of myoglobin. The structure of the illustrated figure was obtained from the 1MBN entry from the protein data bank.[57]

Myoglobin

In papers III–VI, myoglobin was studied in solution with water and trehalose, since it is a well-studied and relatively "simple" protein. Myoglobin was the first protein which structure was determined (via X-ray crystallography in 1958),[58]

and has since then been often used as a model protein. It is a relatively small (∼17.6kDa*) globular protein consisting of eight α-helices and a heme group (see Figure 2.6). The protein stores oxygen in the heme group, specifically in muscle tissue.[56]

2.4.2 Protein water and dynamics

The view of proteins is often simplified as being fixed structures, which functions are only determined by their surface geometry. However, a functioning protein is almost always in an aqueous environment. In fact, the most prevalent molecule by mass in a majority of organisms is water.[59] Protein functions thus occur in the presence of water, and different configurations of a protein are normally slaved to this environment.[60–62] Thus, the understanding how proteins function is not always merely a question of their static geometry, but on their dynamic behavior, and how this is affected by hydration or variation in local environment, by e.g. addition of different solutes. This is of course particularly important when considering the effects of protein stability in different environments.

Water at protein surfaces behaves – just as when water binds to any other type of solute – very different from that in bulk. Hydration water, i.e. water confined by interactions to protein surfaces does not participate in ice formation. Such water has also much slower dynamics than bulk water. For water contents up to about 17 wt% the protein does not exhibit any function, and the water that adsorbs to the protein surface mainly bind to the most hydrophilic sites. As the hydration increases to 30 wt%, the protein is generally hydrated with a single water layer and the function then resumes. The addition of water enables dynamical behavior of the protein; at first, local motions related to e.g. rotational motions of protein side-chains or dipole reorientations give rise to the so called β-relaxations, but as more water is added collective motions –

*For myoglobin obtained from horse heart-muscles

coupled to the α-relaxation – can emerge. Full protein function does however not return until hydration of about 50 wt%, i.e. equal amounts of water and protein,[63] at which point the protein flexibility associated with adding water stops increasing substantially.

There are many different types of protein motions. Below the glass transition temperature (T_g), hydrated proteins exhibit local relaxations, originating from small side-chain fluctuations. At these colder temperatures these motions are thought to be controlled by the local β-relaxation of the solvent. Above T_g, the solvent molecules start to move cooperatively and the β-relaxation of the solvent merges with the more large-scale α-relaxation.[64,65] As the solvent begins to exhibit an α-relaxation, so does the protein molecule, which is what facilitates the protein to undergo large-scale conformational changes, often needed for the protein to function. By understanding the nature of these motions and how they couple to different co-solvents, such as trehalose, one may understand how to modify and improve protein stabilization effects.

2.5 Stabilization of biomolecules

Although the work done in this thesis is focused on the effects of trehalose, this section will first describe why some solutes in general are able to stabilize biomolecules, and secondly, in section 2.5.3 the discussion will focus on what makes trehalose such an excellent protective solute.

2.5.1 Protein stability

Protein dynamics is indeed a criterion for the functionality of proteins, as previously described in section 2.4. However, with great flexibility and motions comes greater risks for protein denaturation to occur. An increase of temperature involves an increase of protein dynamics, which in turn increases the risk for it to irreversibly unfold. Thus, a certain decrease of the dynamics of the

protein is essential for the long-term survivability of protein structures. One method of doing this is to directly slow down the dynamics by decreasing the temperature, so called cryopreservation. This method is common for storing large and complex biological materials, such as cells or tissues, but it is also used for storing proteins for example.[2] Another common method of slowing down the dynamics is to reduce the water concentration, which – as pointed out in section 2.4 – reduces the conformational freedoms of proteins. This can be done by many different methods, one of the most common methods include freeze-drying (or lyophilization). Lyophilization is suitable for storing less complex biological materials, such as proteins. It is done by first freezing the material dissolved in water, followed by keeping the frozen material in a vacuum, in which the water subsequently sublimates. One advantage of storing biological material in a lyophilized state is that this material can be kept at a much higher temperature (e.g. room temperature).[6] There are however problems with both of these two different methods which have to be overcome for improving biological storage techniques.

Cryoprotection

Depressing ice formation is probably the most important aspect of a successful preservation of biological material at low temperatures.* Ice crystals can grow extra- and intra-cellularly puncturing or completely bursting membranes. Another serious problem with ice formation is that it effectively concentrates different chemicals in the cell to high, sometimes toxic, levels. This is due to the fact that ice does not dissolve any substantial amount of solute itself. Similarly, if extracellular water freezes, an effective osmotic stress arises due to the concentration of solutes at the cell membrane. This creates a concentra-

*There are methods that do not involve the avoidance of crystallization, but instead attempts to control the crystallization process by e.g. producing nucleation sites for the crystal structures to grow on (see e.g. Ref. 2). Some organisms use a version of this method by use of antifreeze proteins (see e.g. Ref. 66 for further information). These methods are however beyond the scope of this thesis.

tion gradient that effectively dries out the cells.[67,68] Furthermore, a crystalline ice structure can destroy the native configuration of proteins by forcing these into unfavorable conformations through direct interactions between ice and protein.[69] It is therefore desirable to freeze the protein solution in such a way that it avoids crystallization during cooling, by first entering the supercooled regime and then the glassy state. The glassy state is thus – at least in principle – a way of obtaining a solid but liquid-like environment that very much adapts its structure to that of the protein. Hence the risk for unfavorable steric constraints from a crystal structure is avoided.

Reaching the glassy state and avoiding crystallization is commonly done by the addition of cryoprotectants. These are typically non-reactive molecules that interfere with water crystallization and induce glass formation. There are however some general problems with cryoprotectants. In order to achieve a vitrified state, high concentrations of cryoprotectants are typically required, and that can in itself be toxic. Some cryoprotectants are more toxic than others, but generally, all cryoprotective compounds are damaging at a high enough concentration.[68] Another problem is that the loading of cryoprotectants into a cell may induce osmotic stresses; if a cryoprotectant does not enter through the cell membrane fast enough, the buildup of cryoprotectant extracellularly dries out the cell, possibly leading to a volume collapse. Similarly, after successful cryostorage, the cryoprotectants have to be washed away during reheating, and a buildup of high intracellular cryoprotectant concentration may occur during that stage. This leads to an expansion of the cell, which again may lead to cellular rupture.[68,70]

Although avoiding ice formation may be crucial for successful cryopreservation, it has been shown to not be a sufficient property. Some co-solutes which depress ice formation and increase T_g offer little to no stabilization at low temperatures.[71] So even though crystallization prevention and glass forming abilities seem to be near essential for successful storage of biomolecules,

there has to be some other mechanisms that certain cryoprotectants provide which others lack.

Lyophilisation

Removing water is another method of immobilizing biological molecules. The reduction of water concentration leads to a reduction of molecular dynamics and biomolecular functionality, which has a stabilizing effect. However, when a biomolecule become desiccated, the loss of its function can also have negative effects on its stability.[72] If enough water disappears it could lead to the aggregation and denaturation of protein[73,74] and the destruction of different kinds of cell membranes.[75] This denaturation of protein from either freeze-drying or just air-drying appears to occur when the first hydration shell disappears.[73,74]

In order to obtain a successful freeze-drying, it is also often necessary to add protective molecules to ensure stability at a desiccated stage. It has been pointed out however that the stabilizing mechanism of lyophilization-protectants is fundamentally different from, and much more complex than, that of cryoprotection.[76–78] This aspect can be indirectly evidenced by the fact that most cryoprotective molecules are not capable of desiccation-protection. Molecules that actually have this capability of both cryo- and desiccation-protection are disaccharides.[7,79,80] Disaccharides may thus tell us a bit more about stabilization properties in general than other co-solutes.

2.5.2 Stabilizing co-solutes

A well-proven thermodynamic theory of stabilization/destabilization mechanisms through the addition of solutes has been developed by Timasheff et al. (see for example Ref. 81 or Ref. 82). According to this theory, the stability of the functional state of a protein is proportional to the difference in free energy between the functional state and the denatured state. The greater the difference, the more energy is required to break the functional state.

Generally, stabilizing co-solutes in dilute solutions with protein have been found to be preferentially excluded from the protein surfaces, leading to preferential hydration.[83,84] It was pointed out that a preferentially excluded solute increases the free energy of the material.[83,84] This increase is proportional to the surface area of the protein, and since the surface area is larger for the denatured protein, the free energy of the denatured state will increase more than that of the functional protein state.

To elaborate on that, the stabilization due to the preferential exclusion of co-solutes from the protein surface comes from that if the co-solutes have an unfavorable interaction with the protein surface, the unfolded state will (dependent on the specific protein, but in general) have a larger surface area with more available unfavorable interaction sites. Another way of putting this is that the free energy of the solution increases when the co-solute is added, but it would increase even more if the proteins were to unfold.[82,85]

2.5.3 Bioprotective properties of trehalose

Trehalose emerges in more and more areas as a particularly excellent biological stabilizer.[21–30] Not only as a protectant against cold[23,24] or heat,[21,24–27] but also as a lyophilization stabilizer.[22,28,29] It also has the ability to stabilize both lipid bilayers and proteins.[28] The stabilization of proteins will however be the main focus of this thesis. What exactly it is that makes trehalose different from any other co-solute is not entirely understood, although there are some debated theories, which will be discussed below, and is further investigated in the attached papers.

Vitrification

Stabilizing co-solutes are typically associated with being glass formers. By encapsulating proteins in a glassy matrix, the dynamics of a biomolecule slow down due to the decrease of solvent motions, but without the negative effects

associated with crystallization (e.g. unfavorable molecular geometry). Compared to other similar disaccharides, trehalose has been reported to have the highest glass transition temperature, and to be an excellent glass former. This observation led to the so called *vitrification hypothesis* – as first proposed by Green and Angell[33] – which suggests that the main reason for trehalose's exceptionality resides in its glass forming properties. The reason for this high glass forming property was previously discussed in e.g. section 2.3.3, and is further discussed in papers I–II. Furthermore, when exposed to moisture, trehalose remains at a stable T_g for higher water contents than other disaccharides,[34,35,86] which was discussed in section 2.3.1.

This view – that trehalose's protective properties are due to its extraordinary glass forming properties – was however challenged by a range of studies where some researchers showed that, by using even better glass formers – such as dextran – trehalose was still better at preventing degradation of biomolecules.[87,88] Thus, although the excellent glass forming properties of trehalose may be a very important aspect of its stabilizing properties, it is not sufficient to explain what makes trehalose special. Rather, there ought to be some more intricate interaction between trehalose and biomolecules. Several different models exist that describe this type of interaction, and some of them will be presented here. Specifically the water replacement model, and the preferential hydration model.

Water replacement model

It has been proposed that one important mechanism for further trehalose stabilization stems from direct trehalose–protein interactions. This model presumes that there is a preferential interaction between protein and trehalose, and the hydration layer would then be (at least partially) substituted with trehalose. This is hypothesized to yield a stabilizing effect due to that the protein is kept in its configuration through a direct coupling to a rigid trehalose matrix (see e.g.

Ref. 79). Water replacement with trehalose has been shown to occur in the stabilization of lipid bilayers,[28] however it is less clear if this occurs during protein stabilization. The water replacement model at first showed that the mechanism of stabilization during freeze-drying differed from that of e.g. cryoprotection. The cryoprotection mechanism had been proposed to act according to the preferential hydration model[81–84] mentioned above, but this model was based on studies with relatively high water contents. The water replacement model on the other hand was primarily concerned with the stabilization mechanism at low water contents (e.g. during lyophilization). According to this model, water replacement by trehalose occurs at extremely low water contents. This view was however later disputed by e.g. Belton and Gil et al.,[89] who proposed an alternative theory: the water entrapment model, a variation of the preferential hydration model.

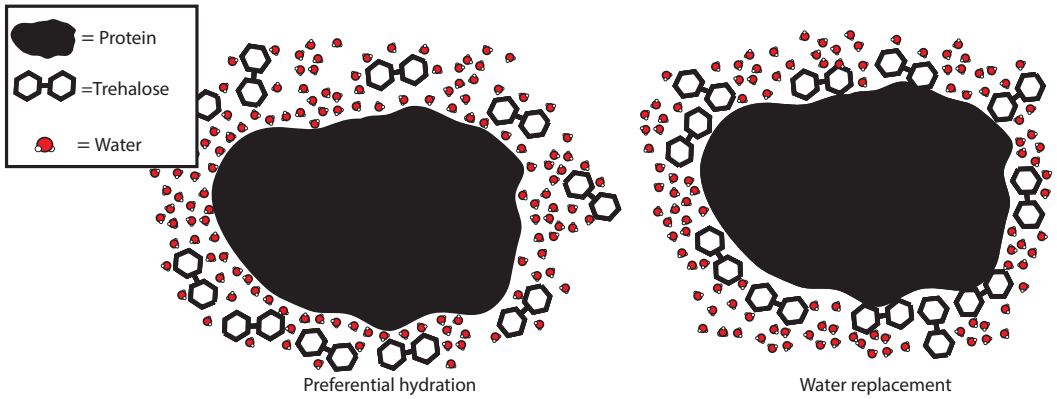


Figure 2.7: Sketch of preferential hydration model (left) and water replacement model (right). According to the preferential hydration model, the native structure of the protein is maintained through direct interaction between water and protein. In the water replacement theory, the hydration shell of the protein is partially replaced by trehalose molecules.

Preferential hydration model

Rather than direct interactions between trehalose and protein, it has been suggested that water molecules prefer to bind directly to the protein surfaces in a solution.[83,84] This effect is highly important, considering that if a layer of

water molecules surrounds the protein, its native solvated state is preserved, even though it may be embedded in a rigid trehalose matrix. The water entrapment theory is a special case of the preferential hydration theory, when the trehalose-protein matrix has a lower water concentration. It has been shown that this interaction remains even when the solution is freeze-dried,[89] indicating that trehalose entraps a hydration layer around the proteins (although some trehalose–protein interaction may become prevalent for cases of extreme desiccation). Support for this model was also found in this work, as presented in papers III – V.

Furthermore, molecular dynamics simulations have shown in extension to this model that in moderately diluted protein-trehalose-water-systems, the trehalose molecules cluster around protein molecules, although without expelling the water molecules.[90,91] Close to the protein, the trehalose molecules and the protein compete over the water molecules, which effectively reduces the strength of the water-protein interaction, thus promoting intramolecular protein-protein interactions which leads to a stabilized protein.[90] If such trehalose structures around proteins are formed it would indicate that trehalose provides a more concentrated stabilizing effect than if the trehalose molecules would have been distributed homogeneously throughout the solution.

Dynamic coupling of proteins with trehalose

Several studies have pointed out that trehalose, compared to its homologous disaccharides, has a strong reducing effect on the dynamics of water.[50,92–94] This effect clearly plays a large role for why trehalose has a relatively high T_g and viscosity in aqueous solutions. However, this strong dynamic coupling could have other important effects for protein stabilization apart from these properties. It has also been suggested – in line with the water entrapment model where the protein is preferentially hydrated – that stabilizing disaccharides mainly couple its dynamics indirectly to the protein *via* the water molecules at the protein

surface, and that trehalose exhibits the strongest dynamic coupling.[95,96] This would make sense considering the aforementioned strong dynamic coupling between trehalose and water.

Other important mechanisms

When some sugar molecules interact with amino acids they may both partly dissociate and form complexes at certain (mostly elevated) temperatures. This reaction is called the Maillard reaction and is the cause for the often desired browning of food, such as toasting of bread. If a sugar and a protein undergo a Maillard reaction, then it follows that there is a subsequent loss in the functionality of the protein.[97,98] It has been shown[99] that trehalose barely exhibits a Maillard reaction with amino acids at all, as opposed to e.g. sucrose which is highly prone to decompose into monosaccharides, which in turn are likely to undergo this type of reaction under heat treatment. It has been proposed that this ability to avoid reactions between sugar and protein is one of the reasons behind the superior protective abilities of trehalose.[25]

2.5.4 Protein aggregation

Protein aggregation is another important issue for the study of protein stability. Proteins may undergo unfolding of their native structure in a reversible manner, without it affecting the long-term function of the protein. However, the probability of said unfolding to be irreversible increases with increased protein-protein interactions.[100,101]

Avoiding protein aggregation is an important field for medical and pharmaceutical purposes. There are many diseases related to proteins aggregating within cells, such as Alzheimer's and Parkinson's disease[102,103]. Likewise, when storing complex pharmaceutics in a medium it is of utmost importance to their viability that they remain as a homogeneous solution.[100] Thus methods of preventing protein aggregation have enormous potential benefits for society. There are nu-

merous reasons for why protein forms aggregates, such as thermal, chemical, or osmotic stress.[104] As previously mentioned, protein aggregation appears to be accompanied by unfolding, or partial unfolding of a protein (see e.g. Ref. 104 and references therein). Thus, protein aggregation can be prevented by adding stabilizing co-solvents, which prevents proteins from unfolding. On the other hand, there have been experiments which show that certain highly unfolded protein structures do not aggregate, and thus there ought to be other effects controlling protein aggregation. Other important mechanisms include the different intermolecular forces between proteins; these forces are governed by the specific geometry and molecular composition of each protein, but also on solution properties, such as pH. If one can tune the solution such that it maximises the protein-protein repulsive forces, the probability of protein aggregation is decreased.

It is therefore hardly surprising that trehalose, with its stabilizing properties, also reduces protein aggregation. This has been specifically shown for a number of studies both *in vivo*[102,103,105–107] and *in vitro*[108–111], and was also shown in paper IV. There it is furthermore hypothesised – based on the aforementioned studies by Lins et al[90] and Corradini et al[91], which suggested that trehalose forms a surrounding layer around the proteins – that surrounding trehalose molecules might prevent proteins from direct interaction by steric hindrance, however no evidence for this hypothesis is given at the time of writing.

2.6 Summary

In summary, there are many different aspects to be concerned about regarding the excellent stabilizing properties of trehalose. In the discussion above, a couple of important hypotheses have been brought up which has been of most relevance to the current thesis; particularly the issue of the structure of aqueous trehalose or the issue of water replacement/entrapment-theories.

There are probably different protective mechanisms for different types of en-

vironmental stresses. For example, the mechanism that stabilizes during cryoprotection is not the same mechanism which is important for lyophilization. However, some factors are very important for a molecule to be able to stabilize biological materials. It needs to be able to slow down the surrounding water dynamics and perturb the water structure, yielding a stabile glass without crystallization of the water. It should also be non-reactive with the protein, and be able to protect many different types of biomolecules (such as both proteins and lipid bilayers). Trehalose seems to possess all these qualities. It has a strong effect on water and it easily perturbs crystallization and enables glass formation. Moreover, it prefers to interact with water over protein, thus leaving the protein hydrated, although with a more rigid hydration shell. The interaction between trehalose and the hydration layer of the protein also creates weaker protein-water bonds, which effectively strengthens intramolecular protein interactions. Trehalose also appears to prevent proteins from aggregating, however the exact nature of this effect remains to be investigated.

3

Experimental and Computational Methods

The main method used in this thesis is neutron diffraction combined with empirical potential structure refinement (EPSR) modeling, used for the investigation of the structure of aqueous sugar and/or protein samples as in papers I, II, and IV. Furthermore, differential scanning calorimetry (DSC) was used for indirect structural investigations of aqueous trehalose containing protein (paper III), quasielastic neutron scattering (QENS) in combination with molecular dynamic (MD) simulations was used in paper V to investigate the dynamics of aqueous samples of either trehalose, myoglobin, or both. Finally, broadband dielectric spectroscopy (BDS) was used for the investigation of different relaxation phenomena of relatively dry myoglobin/trehalose/water samples in paper VI. These six methods will be the main focus of the following chapter, containing a brief overview of the theory behind these techniques. More details about the implementations of these techniques will be further discussed in chapter 4.

3.1 Neutron Scattering

The use of scattering techniques for the study of all types of materials (including biological material) have been around for a long time. The first images of a protein (myoglobin) for example was, as previously mentioned, produced by Kendrew et.al., using X-ray scattering.[58] Scattering methods in general has grown a lot in use during the past decades. For the investigation of biological materials (and many other kinds of materials) there are often a lot to be gained from using neutrons rather than X-rays, since the neutrons are electrically neutral and are capable of penetrating deep into the material. However, perhaps the biggest advantage of using neutrons is that hydrogen – which is extremely abundant in biological materials – has a high cross section for neutron interaction, and is thus easier to study using neutrons. The cross section of atoms using X-rays is proportional to the number of electrons in each atom, and thus hydrogens are objects which are difficult to detect. The cross section of atoms as seen by neutrons however, vary more sporadically for different elements and isotopes. Furthermore, since the neutrons interact with atomic nuclei, this technique offers itself to the possibility of performing isotope substitutions. By substituting e.g. hydrogen with deuterium, which have very different total cross sections, it is possible to alter the contrast between atoms depending on its isotope composition (see also section 4.1.3), without altering the structural and dynamical properties of the material substantially.

A typical neutron scattering experiment starts by irradiation of a sample with a beam of neutrons. The neutrons are either produced in a reactor – where the fast neutrons are the by-product of radioactive decay from uranium (^{235}U) – or by a spallation source, in which high energetic protons are collided with a block of tungsten which then eject ("spall") neutrons in every direction. From a spallation source the neutrons are guided through e.g. moderators (decreasing the kinetic energy of the neutrons), collimators, monochromators, or other

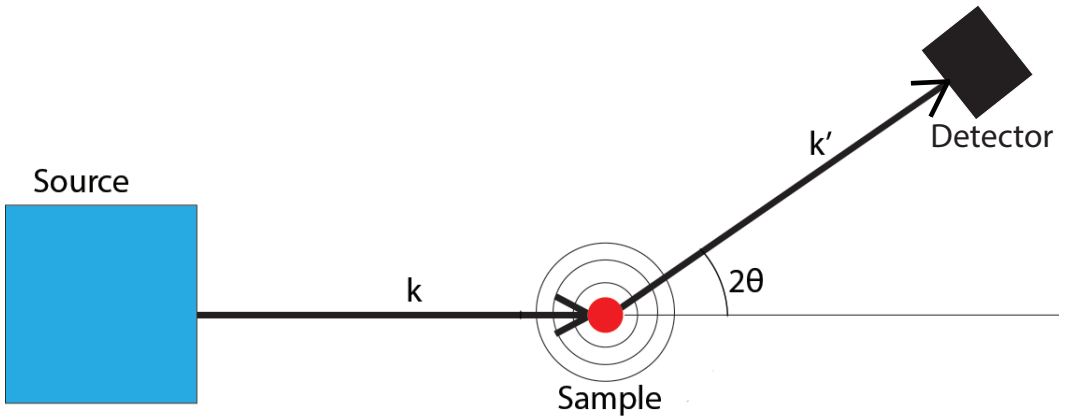


Figure 3.1: Schematic of a neutron scattering event. Incident neutrons impinge on the sample as a plane wave, with wave-vector k, and scatter radially with a wave-vector k'.

devices designed to select the desired properties of the neutrons destined to hit the sample. Some of the neutrons that hit the sample scatter and are subsequently detected, and data are collected for further analysis.

There are many types of different neutron scattering techniques, which detect or focus on different properties of the scattered neutrons. In the present work, neutron diffraction experiments, and quasi-elastic neutron scattering (QENS) have been used. Neutron diffraction methods mainly yield information about the structure of the investigated materials, and QENS yields information about atomic dynamics for certain time-intervals.

3.1.1 Neutron Diffraction

Neutron diffraction is an experimental technique where one focus on the elastic coherent part of the scattering. The scattered neutrons mainly contain information about structural correlation lengths within the sample. In the case of a neutron diffraction experiment, such as the one performed in this work (performed on NIMROD, see section 4.1.1 for further details), the sample is hit by a pulse of neutrons with wave-vectors $\mathbf{k}$, with neutron wavelengths of λ_i (ranging from 0.05 AA to 10 Å), and scatter with wave-vectors $\mathbf{k}'$. In a neutron

scattering experiment, the sample chamber is surrounded by detectors which essentially counts the number of neutrons scattered at different angles, and also at different arrival times. The scattered neutron wavelengths at a particular angle is calculated from their times of flight. Thus, the raw data produced is an intensity distribution function, $I(2\theta, \lambda)$, that relates the neutron intensity to a specific angle and wavelength. Typically, $I(2\theta, \lambda)$, is written as a function $I(\mathbf{Q})$, where $\mathbf{Q}$ is the scattering vector defined as*:

$$\mathbf{Q} = \mathbf{k} - \mathbf{k}' \tag{3.1}$$

where the relationship between $|\mathbf{Q}| = Q$ and θ can be found by approximating the scattering to only be elastic scattering, such that $|\mathbf{k}| = |\mathbf{k}'| = \frac{2\pi}{\lambda_i}$. From that approximation and simple trigonometry one can obtain:

$$Q = \frac{4\pi}{\lambda} \sin(\theta) \tag{3.2}$$

So how does this scattered intensity relates to the structure of the material? Recall section 2.1.2 where the pair correlation function $g(r)$ was introduced, which is perhaps the simplest way to represent the structure of an amorphous material. In the remaining part of this section a brief outline is given on how to get from the measured diffracted neutron data to $g(r)$.

First of all the obtained raw data from the diffraction experiment have to be corrected for different background signals, neutron absorption events, etc., which is given an overview of in section 4.1.2. From those corrections, the double differential scattering cross section $\frac{d^2\sigma}{d\Omega dE}(Q)$ is obtained, which is the number of neutrons scattered into a solid angle element $d\Omega$ at an angle corresponding to scattering vector Q, with an energy between E and $E + dE$. When studying elastic scattering, the energy part is integrated out, according

*The derivation of the formalism in this section can be found in multiple textbooks. For more details the reader is referred to e.g. Ref. 112–115

to equation 3.3, and the double differential scattering cross section can thus be reduced to the differential scattering cross section $\frac{d\sigma}{d\Omega}(Q)$.

$$\frac{d\sigma}{d\Omega} = \int_{-\infty}^{\infty} \frac{d^2\sigma}{d\Omega dE}(Q)dE \tag{3.3}$$

More about how to account for inelastic effects can be found in e.g. Ref. 116 or 117, however this integration is valid for the present discussion which will focus on the static differential scattering cross section: $\frac{d\sigma}{d\Omega}(Q)$.

So what is then the origin of this differential scattering cross section? Let us consider the wave-formalism for neutrons for a while*; the neutron beam impinges on the sample as a plane wave (collimated beam) with wave-vector $\mathbf{k}$, and can be written as $\psi_i = \psi_0 exp(i\mathbf{k}\cdot\mathbf{r})$ at position $\mathbf{r}$. ψ_0 is the amplitude of the incident wave related to the flux of the beam. After interacting with a single nucleus labeled j at position $\mathbf{r}_j$, the outgoing wave will propagate radially outwards from the nucleus as:

$$\psi_f = -\psi_0 e^{i\mathbf{k}\cdot\mathbf{r}_j} b_j \frac{e^{i\mathbf{k}'\cdot(\mathbf{r}-\mathbf{r}_j)}}{|\mathbf{r}-\mathbf{r}_j|} \tag{3.4}$$

where b is the so called scattering length of the nucleus, related to the cross section ($\sigma = 4\pi|b|^2$, measured in barns) of the nucleus, and is a measure of how strong the interaction between the incident neutron and the nuclei is. Summing up the contributions from all N atoms in the sample, one obtains the total wave function:

$$\psi_f = -\psi_0 e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{j=1}^{N} b_j \frac{e^{i\mathbf{Q}\cdot\mathbf{r}_j}}{|\mathbf{r}-\mathbf{r}_j|} \tag{3.5}$$

where again, $\mathbf{Q} = \mathbf{k} - \mathbf{k}'$. This wave function describes the amplitude of the scattered wave at different positions. The position of each atom could in principle be obtained through a Fourier transform of this function. However, what is measured at the detector is not the wave itself, but the square modulus

*For simplicity, the formalism described here applies for single component systems.

of the wave function, $|\psi_f|^2$.

$$|\psi_f|^2 \approx \frac{|\psi_0|^2}{r^2} \sum_{i=1}^{N} b_i e^{i\mathbf{Q}\cdot\mathbf{r}_i} \cdot \sum_{j=1}^{N} b_j^* e^{-i\mathbf{Q}\cdot\mathbf{r}_j} \tag{3.6}$$

In equation 3.6, it was assumed that the distance between the detector and the sample is much greater than the distances between nuclei in the sample ($|\mathbf{r}| >> |\mathbf{r_j}|$), and thus $|\mathbf{r} - \mathbf{r}_j| \approx r$. This equation can furthermore be rewritten in terms of the differential scattering cross section by noting that the fraction of the incident neutrons which impinges on a small area, dA, is $d\sigma = \frac{|\psi_f|^2}{|\psi_0|^2} dA$. The differential scattering cross section can thereby be written as the fraction of incident neutrons per solid angle $d\Omega$:

$$\frac{d\sigma}{d\Omega} = \frac{|\psi_f|^2}{|\psi_0|^2}\frac{dA}{d\Omega} = \left\{ d\Omega = \frac{dA}{r^2} \right\} = \sum_{i=1}^{N} b_i e^{i\mathbf{Q}\cdot\mathbf{r}_i} \cdot \sum_{j=1}^{N} b_j^* e^{-i\mathbf{Q}\cdot\mathbf{r}_j} \tag{3.7}$$

Rewriting equation 3.7 and averaging the nuclear scattering length, b, over different spin orientations and isotopes the following equation can be obtained:[115]

$$\frac{d\sigma}{d\Omega} = \sum_{i,j}^{N} \langle b_i b_j^* \rangle e^{i\mathbf{Q}\cdot(\mathbf{r}_i - \mathbf{r}_j)} \tag{3.8}$$

It is useful to separate the term $\langle b_i b_j^* \rangle$ into two different cases as well, either $i \neq j$ or $i = j$. If $i \neq j$ then the average is taken over two different atoms which do not have any correlation in their scattering lengths. Hence for $i \neq j$ then $\langle b_i b_j^* \rangle = \langle b_i \rangle \langle b_j^* \rangle = \langle b \rangle^2$. In the case when $i = j$ however, the scattering refers to the "self-scattering" of individual nuclei. The term $\langle b_i b_j^* \rangle$ then becomes: $\langle b^2 \rangle = \langle b \rangle^2 + \left\langle (b - \langle b \rangle)^2 \right\rangle$, i.e. a measure of how much the scattering length deviates from the mean value.[112]

Using these scattering lengths and that for $i = j$ then $r_i = r_j$, then equation

3.8 can be written as:

$$\frac{d\sigma}{d\Omega} = \underbrace{\langle b\rangle^2 \sum_{i,j}^{N} e^{i\mathbf{Q}\cdot(\mathbf{r}_i-\mathbf{r}_j)}}_{\text{Coherent scattering}} + \underbrace{N(\langle b^2\rangle - \langle b\rangle)^2)}_{\text{Incoherent scattering}} \tag{3.9}$$

The incoherent scattering part is clearly not dependent on any structural parameters concerning the sample, and is thus merely added to the total scattering cross section as an (often unwanted) background in the case of a diffraction study. The coherent part however is typically written as:

$$\frac{d\sigma}{d\Omega}_{coh} = \langle b\rangle^2 N S_{coh}(\mathbf{Q}) \tag{3.10}$$

Where $S_{coh}(\mathbf{Q}) = \frac{1}{N}\sum_{i,j}^{N} e^{i\mathbf{Q}\cdot(\mathbf{r}_i-\mathbf{r}_j)}$ is the coherent structure factor (from now on just called the structure factor, or $S(\mathbf{Q})$), which is a function describing the system that only depends on the investigated material. Using a general property of the Dirac delta function that $\sum_i e^{i\mathbf{Q}\cdot\mathbf{r}_i} = \int_V e^{i\mathbf{Q}\cdot\mathbf{r}_i}\sum_i \delta(\mathbf{r}-\mathbf{r}_i)$, $S(\mathbf{Q})$ can be rewritten as:

$$S(\mathbf{Q}) = 1 + \frac{1}{N}\int e^{-i\mathbf{Q}\cdot\mathbf{r}} \sum_{i,j\neq i}^{N} \delta(\mathbf{r} - (\mathbf{r}_i - \mathbf{r}_j))d\mathbf{r} \tag{3.11}$$

The summation of Dirac delta functions in equation 3.11 is familiar from section 2.1.2. By recalling the definition of $g(r)$ through equation 2.3 and inserting this into equation 3.11, the following equation is obtained:[118]

$$S(\mathbf{Q}) = 1 + \rho\int e^{-i\mathbf{Q}\cdot\mathbf{r}} g(\mathbf{r})d\mathbf{r} \tag{3.12}$$

which, for an isotropic liquid can be simplified to:

$$S(Q) = 1 + 4\pi\rho\int_0^\infty r^2 g(r)\frac{\sin(Qr)}{Qr}dr \tag{3.13}$$

It can thus be concluded that there exist a relationship between the obtained neutron diffraction data and the pair correlation function.

Partial structure factors

The structure factor $S(Q)$ describes the total coherent elastic scattering of the sample. However, it is sometimes useful to divide $S(Q)$ into a sum of contributions arising from the correlations between different atom types. Similar to what was done for the total pair correlation function into partial pair correlation functions in section 2.1.2. The individual terms in $S(Q)$ is called partial structure factors and are usually written as $S_{\alpha\beta}(Q)$, where α and β represent different atoms. They are defined by:

$$S(Q) = \frac{\sum_{\alpha\beta}\langle b_\alpha\rangle\langle b_\beta\rangle(S_{\alpha\beta}(Q)-1)}{\sum_\alpha c_\alpha\langle b_\alpha\rangle^2} + 1 \tag{3.14}$$

Where again, c_α and c_β are the fractions of atoms α and β in the sample. The partial structure factors in turn relate to the correlations between atom types α and β via the partial pair correlation functions $g_{\alpha\beta}$ through:

$$S_{\alpha\beta}(Q) = c_\alpha\delta_{\alpha\beta} + c_\alpha c_\beta\rho\int e^{-i\mathbf{Q}\cdot\mathbf{r}}g_{\alpha\beta}(r)d\mathbf{r} \tag{3.15}$$

More details about how these partial pair correlation functions are found based on the obtained $S(Q)$ will be treated in section 3.2.

3.1.2 Quasielastic Neutron Scattering

In the previous section, elastic scattering was assumed. This means that the magnitude of the neutron wave-vector remained the same before and after scattering ($|\mathbf{k}| = |\mathbf{k}'|$), and from that equation 3.2 was derived. This is however not the actual case for most samples; if atoms e.g. has some translational motion during scattering, some of that translational energy is transferred to or from

the neutron.* The resulting energy transfer ΔE is related to a frequency shift of the neutron according to the de Broglie equation, $\Delta E = \frac{\hbar^2}{2m}(k^2 - k'^2) = \hbar\Delta\omega$. For the purpose of the diffraction experiments, these energy transfers could be ignored (by performing some data corrections) and the atoms were assumed to be static. However, for the purpose of studying the dynamical properties of the investigated material these energy transfers could be measured, thus giving direct information about certain atomic motions. In a QENS measurement one is interested in such inelastic events, for relatively low energy transfers,† corresponding to dynamics such as molecular rotations and translations.

The theory behind QENS can be derived much in the same way as in the elastic case (explained in the above section), but by adding a time-dependency on the positions of the atoms and allowing for energy transfer, the measured signal now becomes a double differential scattering cross section $\frac{d^2\sigma}{d\Omega dE}$,‡ which turns out similar to that of equation 3.8:

$$\frac{d^2\sigma}{d\Omega dE} = \frac{1}{\hbar}\frac{d^2\sigma}{d\Omega d\omega} = \frac{k'}{k}\frac{1}{2\pi}\int_0^\infty e^{-i\omega t}\sum_{i,j}^{N}\langle b_i b_j^*\rangle e^{i\mathbf{Q}\cdot(\mathbf{r}_i(0)-\mathbf{r}_j(t))}dt \tag{3.16}$$

Where, the term $\frac{k'}{k}$ is the ratio between the amplitude of the scattered and the incident wave-vectors. Just like in the previous section for equation 3.8, equation 3.16 can be divided into inter-atomic scattering, and self-scattering, corresponding to the coherent and incoherent scattering respectively:

*The following equation is instead used to calculate the scattering vector taken energy transfer into account[115]: $Q = \sqrt{k^2 + k'^2 - 2kk'\cos(2\theta)} = \frac{8\pi^2}{\lambda^2} + \frac{2m\omega}{\hbar} - \frac{4\pi}{\lambda}\sqrt{\frac{4\pi^2}{\lambda^2} + \frac{2m\omega}{\hbar}\cos(2\theta)}$

†Relatively low energies corresponds to relatively slow atomic motions, i.e. "almost" static, hence the term "quasi" in quasielastic neutron scattering

‡Fraction of incident neutrons per solid angle $d\Omega$ with final energies of E and E+dE

$$\frac{d^2\sigma}{d\Omega dE} = \underbrace{\frac{k'}{k}\frac{1}{2\pi}\langle b\rangle^2 \int_0^\infty e^{-i\omega t}\sum_{i,j\neq j}^{N} e^{i\mathbf{Q}\cdot(\mathbf{r}_i(0)-\mathbf{r}_j(t))}dt}_{\text{Coherent scattering}} + \underbrace{\frac{k'}{k}\frac{1}{2\pi}(\langle b^2\rangle - \langle b\rangle)^2)\int_0^\infty e^{-i\omega t}\sum_{i}^{N} e^{i\mathbf{Q}\cdot(\mathbf{r}_i(0)-\mathbf{r}_i(t))}dt}_{\text{Incoherent scattering}} \quad (3.17)$$

This can be rewritten in terms of the coherent and incoherent dynamic structure factors, $S_{coh}(\mathbf{Q},\omega)$ and $S_{inc}(\mathbf{Q},\omega)$, defined as:

$$\frac{d^2\sigma_{coh}}{d\Omega dE} = \frac{k'}{k}\frac{N}{2\pi}\langle b\rangle^2 S_{coh}(\mathbf{Q},\omega) \quad (3.18)$$

$$\frac{d^2\sigma_{inc}}{d\Omega dE} = \frac{k'}{k}\frac{N}{2\pi}(\langle b^2\rangle - \langle b\rangle)^2)\ S_{inc}(\mathbf{Q},\omega) \quad (3.19)$$

Where $S_{coh}(\mathbf{Q},\omega)$ and $S_{inc}(\mathbf{Q},\omega)$ are properties that are only dependent on the structure and dynamics of the sample. These two dynamic structure factors can furthermore be rearranged using the general property of the Dirac delta function as used for equation 3.11, and the van Hove equations 2.6 and 2.7:

$$S_{coh}(\mathbf{Q},\omega) = \int_0^\infty \sum_i^N e^{i(\mathbf{Q}\cdot(\mathbf{r}_i(0)-\mathbf{r}_j(t))-\omega t)}dt = \\ = \int_0^\infty \int_{-\infty}^\infty e^{i(\mathbf{Q}\cdot(\mathbf{r}_i(0)-\mathbf{r}_j(t))-\omega t)} G_{Distinct}(\mathbf{r},t)d\mathbf{r}dt \quad (3.20)$$

$$S_{inc}(\mathbf{Q},\omega) = \int_0^\infty \sum_i^N e^{i(\mathbf{Q}\cdot(\mathbf{r}_i(0)-\mathbf{r}_i(t))-\omega t)}dt = \\ = \int_0^\infty \int_{-\infty}^\infty e^{i(\mathbf{Q}\cdot(\mathbf{r}_i(0)-\mathbf{r}_i(t))-\omega t)} G_{Self}(\mathbf{r},t)d\mathbf{r}dt \quad (3.21)$$

Or to put equations 3.20 and 3.21 differently: the coherent and incoherent

dynamic structure factor ($S_{coh}(\mathbf{Q},\omega)$ and $S_{inc}(\mathbf{Q},\omega)$) of a sample is related to the dynamic pair-correlation function and the dynamic self-correlation function respectively ($G_{Distinct}(\mathbf{r},t)$) and $G_{Self}(\mathbf{r},t)$) via Fourier transformation. Thus, incoherent inelastic scattering comes from the self-scattering of an atom at different times, and coherent inelastic scattering comes from relative motions of atoms over time.

QENS analysis

The signal from a typical QENS experiment is often represented as $S(Q,\omega)$, as shown in figure 3.2a for $Q = 1$ Å^{-1}. The midpoint ($\omega = 0$eV), represents the elastic scattering, and the broadening of that signal arises from inelastic contributions where the neutrons have lost or gained energy. Thus, in principle, the more the peak is broadened the more inelastic scattering is present in the sample. However, for the most part of the work in this thesis, the QENS signal has mainly been analyzed from the intermediate scattering signal ($I(Q,t)$), and hence the discussion will focus on this representation. $I(Q,t)$ is obtained by Fourier transforming $S(Q,\omega)$ in the time domain, and the resulting $I(Q,t)$-signal for $Q \approx 1$Å^{-1} can be seen in figure 3.2b as a function of time. Analogous to the interpretation about the broadening of $S(Q,\omega)$, inelastic contributions to $I(Q,t)$ can be seen as a decrease in the signal. Or in other words: an atom's position correlates less with its original position over time if it moves (in the case of incoherent scattering). Thus, if the scattering only originated from fixed atoms (as in a crystal structure), the atoms would fully correlate with themselves and the signal would have a constant value of one.

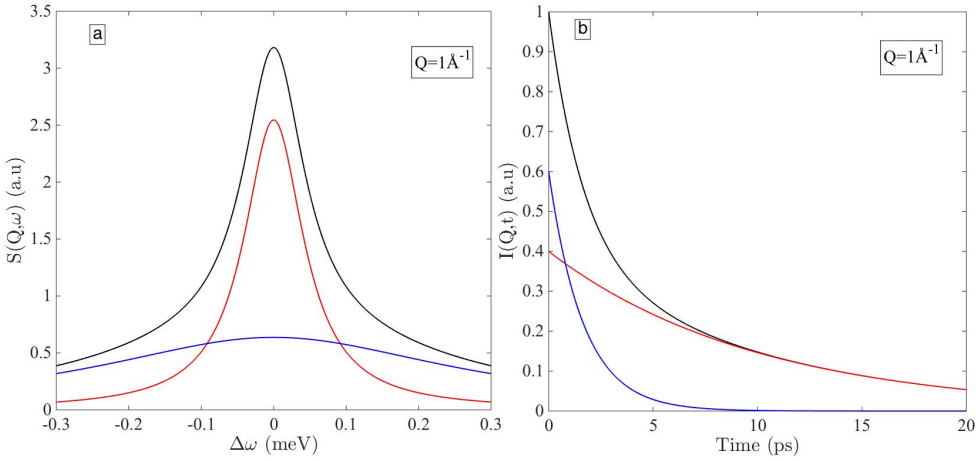


Figure 3.2: a) Dynamic scattering function and b) intermediate scattering function for $Q = 1$ Å^{-1} shown as black curves which is a sum of two different types of dynamics: the red curves indicating a slower relaxation and the blue curves showing a faster relaxation.

The dynamics of an atom type can be described by multiple different models, for example by the use of a stretched exponential function:

$$I_k(Q,t)|_{Q=Const.} = A_k \cdot exp(-\left(\frac{t}{\tau_k}\right)^{\beta_k}) \tag{3.22}$$

Where A_k is the scattering contribution from atom type k to the total scattering, τ_k is the typical relaxation time of that atom type. β_k is a stretching parameter between 0 and 1 where 1 means no stretching, which implies that all atoms of that type have exactly the same relaxation time, and lower values of β_k means that the atoms have a distribution of relaxation times with the typical relaxation time τ_k.

One important method of extracting which relaxation times belong to a specific atom type is isotope substitution of hydrogens to deuterium. Hydrogens have a much larger total scattering cross section than deuterium (82 and 7.6 barns respectively), and thus hydrogens contribute much more to the scattering signal. Thus, in a hydrated protein sample for example, it is possible to "extinct" the contribution of the hydration water by the use of deuterated water, which

means that the internally bound protein hydrogens will dominate the scattering signal.

3.2 Empirical Potential Structure Refinement Modeling

In this section, a couple of methods of obtaining partial pair correlation functions from neutron diffraction data (see section 3.1.1) will be presented. In principle, a full set of partial pair correlation functions of a material can be obtained by isotopically labelling every unique atom type in the material, exchanging each atom type with a different isotope at the time. However, this method would be very time consuming, and such accurate partial isotopical marking is rarely available. Instead, several methods of computer modeling have been developed in order to obtain a complete model of a sample from the available diffraction data.

One such method, used in the work of this thesis, is the EPSR method. To describe this method it is useful to start with a previous method which EPSR is based on, namely the Reverse Monte Carlo (RMC) method.

3.2.1 RMC

In the RMC method (described in more detail in Ref. 119) the researchers typically set up a system where the structure of the molecules is initially defined, and then a hard-sphere model is applied for the interactions between the molecules. A set of constraints are typically also set up, such as minimum intermolecular distances, or specific bond angles, as determined from other experimental methods (e.g. NMR, MD). Such constraints are necessary for avoiding unreliable outcomes, by for example preventing atomic overlaps.
The RMC program then randomly moves around an atom or a molecule and calculates e.g. the structure factors, and compare them with the experimental

structure factors. If the move decreased the difference between the two corresponding structure factors, the move is accepted. If it increased the difference, the move is only accepted with a probability of:

$$exp\Big(- (\chi^2_{before\ move} - \chi^2_{after\ move})/2\Big) \tag{3.23}$$

where χ^2 is a quantity of the difference between the simulated and experimental data, and defined as:

$$\chi^2 = \sum_{i=1}[S_{calc}(Q_i) - S_{exp}(Q_i)]^2/\sigma^2(Q_i) \tag{3.24}$$

where $\sigma(Q)$ is the standard deviation of experimental error for any measured value of Q.

RMC modeling is a widely used method which has been used for decades, and has helped to elucidate the molecular structures of a significant number of disordered materials.[120] However, RMC fails to take into account several physical aspects of a material when just searching for a structure which fits the experimental data, such as the fact that the obtained structure could in principle have a molecular potential which is energetically unfavorable.[119]

3.2.2 EPSR

The EPSR method used in this work is derived from the RMC method, and just like RMC, EPSR seeks to obtain a minimum difference between the derived data and a set of experimental data. This is opposed to the Metropolis Monte Carlo (MMC) method which is an efficient method of obtaining the minimum intermolecular potential energy of a system. However, there are some important differences: in EPSR, the goal is to obtain a correct intermolecular potential which leads to convergence with the experimental data, rather than fitting the simulated structure directly to the diffraction data.

The procedure works in principle according to the following scheme:[121]

1. Create a simulation box with the correct density and intramolecular structure.

2. Assign a reference potential (RP) to the system, i.e. find Lennard-Jones- and Coloumb-parameters for each atom type.

3. Run a MMC simulation on the system to minimize the potential energy. This is done by randomly moving (translation, rotation, or bending) an atom or molecule and subsequently measuring the potential difference (ΔU) due to that move. The move is always accepted if $\Delta U < 0$, and only accepted with a probability of $exp[-\frac{\Delta U}{k_B T}]$ if $\Delta U > 0$.

4. From the energetically minimized structure, the pair correlation function is calculated. This is subsequently Fourier transformed to give the simulated structure factor, $S_{calc}(Q)$.

5. The difference, $S_{calc}(Q) - S_{exp}(Q)$ is then calculated and used to calculate the empirical potential (EP).

6. The empirical potential is then added to the reference potential.

7. Steps 3 - 6 are then iterated until the EP becomes stable, or until its absolute energy exceeds a predefined value.

To start an EPSR simulation, first one needs to find a reference potential. This includes different Lennard-Jones- and Coloumb- parameters but also intramolecular structures, intermolecular starting configurations and minimal distances. Typically, one has a lot of knowledge about a material determined through previous experiments which can be effectively included in the RP. The Lennard-Jones (LJ) parameters sets up an often used potential (Lennard-Jones potential) between two atoms and is the basis of the reference potential combined with an added Coloumb potential:

$$U_{\alpha,\beta}^{LJ} = 4\epsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{r}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r}\right)^{6}\right] + \frac{q_\alpha q_\beta}{4\pi\epsilon_0 r} \tag{3.25}$$

Where r is the distance between atom α and atom β, q is the electric charge of a specific atom, $\sigma_{\alpha\beta}$ is the distance where the potential is zero, and ϵ is the depth of the potential well. $\sigma_{\alpha\beta}$ and $\epsilon_{\alpha\beta}$ are calculated according to the Lorentz-Berthelot mixing rules, based on the LJ parameters of the individual atoms.

$$\epsilon_{\alpha\beta} = \sqrt{\epsilon_\alpha \epsilon_\beta} \qquad \sigma_{\alpha\beta} = \frac{\sigma_\alpha + \sigma_\beta}{2} \tag{3.26}$$

The individual LJ parameters in turn can be determined through various force-field calculations such as the OPLS-AA (Optimized Potentials for Liquid Simulations - All Atoms) force field which was used in papers I, II and IV.

Further additions to the reference potential are well described in the EPSR manual (Ref. 121) and Ref. 120.

Below is given an overview of how the empirical potential (EP) is defined. For a more detailed description about these steps, the reader is again referred to Ref. 121 and Ref. 120. The empirical potential in real space is defined as a sum of Poisson functions:

$$U(r)^{EP} = kT \sum_i C_i p_{n_i}(r, \sigma) \tag{3.27}$$

where

$$p_{n_i}(r, \sigma) = \frac{1}{4\pi\sigma^3(n+2)} \left(\frac{r}{\sigma}\right)^n e^{-\frac{r}{\sigma}} \tag{3.28}$$

and $n_i = \frac{r_i}{\sigma} - 3$ (r_i and σ are set by the user). C_i:s are weights which are fitted through comparison with real data. $p_{n_i}(r, \sigma)$ can be Fourier-transformed to:

$$P_n(Q, \sigma) = \frac{1}{\sqrt{1 + Q^2\sigma^2}^{(n+4)}(n+2)} \left[2\cos(n\alpha) + \frac{(1 - Q^2\sigma^2)}{Q\sigma} \sin(n\alpha)\right] \tag{3.29}$$

where $\alpha = \arctan(Q\sigma)$. The empirical potential can then be written in

Q-space accordingly:

$$U(Q)^{EP} = \sum_i C_i P_{n_i}(r, \sigma_Q) \tag{3.30}$$

The C_i-weights are determined by fitting $U(Q)^{EP}$ to the difference between $S_{exp}(Q)$ and $S_{calc}(Q)$. Once these go to small values (ideally to zero), the EP converges and thus a final total potential is obtained. From this stage it is possible to start analyzing the obtained model and extract useful structural information.

3.3 Molecular Dynamic Simulations

Simulation methods such as RMC and EPSR are excellent when attempting to find a structure corresponding to a particular structural measurement. However, the actual atomic moves these methods produce are stochastic and say little about the true dynamics of the sample. On the other hand, the forces between all atom pairs in a simulation box could however be calculated to a high degree of precision (by e.g. *ab initio* methods). Thus, a theoretical model containing extensive information about both the structure and dynamics of a molecular system can be obtained. This is however not always an appropriate method. Such a thorough calculation over a sufficient number of atoms, and over a satisfyingly long time span demands vast computation power. For this reason, a number of methods assume certain approximations to shorten computational demands. One such powerful method is classical molecular dynamics (MD) simulations.*A brief introduction of the basic principle of a classical MD simulation will be presented in this section.

The first step of an MD simulation is to build a molecular simulation box. This can in principle be a challenging task depending on the complexity of

*MD simulations could refer to many different techniques, but for the purpose of this section it will refer to all-atomic classical MD simulations as used in standard MD simulation software, such as Gromacs[122].

the molecules of interest, and several different software packages have been developed especially for this purpose (e.g. Packmol, CHARMM-GUI, etc.), but for the present discussion this will not be discussed in more detail. To illustrate one such simulation box (used in paper V), figure 3.3 shows a molecular box containing 2000 water molecules and 104 trehalose molecules. Secondly, in a

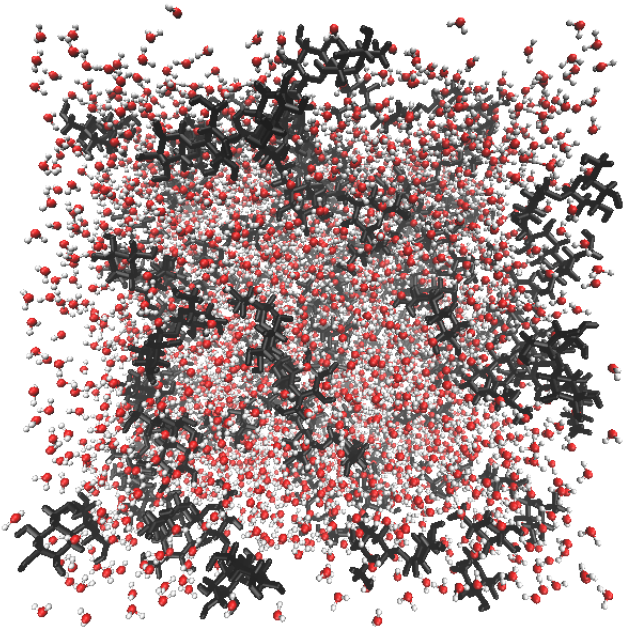


Figure 3.3: Simulation box of trehalose in water. Trehalose is shown as black molecules, and water as red and white molecules.

real sample, all molecules have a certain momentum which is directly related to the temperature of the sample. The initial momentum of the different molecules are randomly assigned such that the average energy of the sample corresponds to the desired temperature of the sample.

Perhaps the most important step in this simulation is now to decide what will happen to the positions of each atom as the simulation progresses over time. This is determined for each specific atom (labelled i) by solving Newton's equation for that atom:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i = -\nabla U \tag{3.31}$$

where m_i and r_i are the weight and position of atom i and $\mathbf{F}_i$ is the force on

atom i, which is determined by the derivative of the potential energy (U) at position r_i. For a given potential energy equation 3.31 can be solved numerically in different ways, such as the Verlet algorithm, or the leapfrog algorithm. To run an MD simulation, one requires information about the potential energy between pairs of atoms, or in other terms: the force field. There are many different force fields commonly used with different advantages and disadvantages regarding accuracy and speed. Examples of such force fields which have been used in the scope of this thesis include the CHARMM (Chemistry at HARvard Molecular Mechanics) force field[123], or the OPLS (Optimized Potential for Liquid Simulations) force field[124].

A common way of defining a force field is via the derivative of a sum of some interatomic potentials. The total potential is a sum of the intramolecular potential (e.g. bond-length or bond-angle energy-contributions), and the intermolecular potential (from e.g. van der Waal- or Coloumb-forces). An interatomic potential was previously mentioned in the EPSR section as the reference potential used in EPSR (equation 3.25). More explicitly, the total potential can be written as:

$$U_i = \sum k_i^{bond}(r_i - r_0)^2 + \sum k_i^{angle}(\theta_i - \theta_0)^2 + \sum k_i^{dihedral}(1 + \cos(n_i\phi_i + \delta_i)) + \sum_i \sum_{j \neq i} 4\epsilon_{i,j}\left[\left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{4\pi\epsilon_0 r_{i,j}} \quad (3.32)$$

Where the first three terms correspond to the potential energy of bonded atoms. The first term gives the potential energy contribution stemming from the displacement of an atom in relationship to its bonded neighboring atoms. It has an equilibrium position (r_0), and a specific bond-strength (k_i^{bond}) describing how much energy it takes to stretch atom i away from its equilibrium position. Similarly for the second term, bonds have a preferred angle within a molecule (θ_0), with a corresponding strength (k_i^{angle}). The third term applies for molecules larger than four atoms, where torsions of different groups or planes of atoms

are important. For a specific dihedral plane, ϕ_i is the dihedral angle, δ_i is the phase, n_i is the number of potential minima the dihedral angle can reside in, and $k_i^{dihedral}$ gives the strength of the potential-well for the dihedral conformation. The two last terms are from non-bonded interactions. The fourth term is the so called Lennard-Jones potential, which is an approximation of the forces due to a combination of the repulsive effect (the r^{-12}-term) due to the Pauli exclusion principle, and the attractive van der Waals force (the r^{-6}-term). $\sigma_{i,j}$ is the equilibrium distance of two atoms i and j, and $\epsilon_{i,j}$ is the depth of the the potential energy-well at that distance. The final term comes from Coloumb-interactions between two atoms with charges q_i and q_j separated with a distance of $r_{i,j}$

Given the initial conditions and the force-field, the system is typically updated with a time-step of typically around 1-2 fs, which is sufficiently fast to capture the fastest dynamics of a typical system. From these time-steps every atomic position and velocity can be updated accurately and saved to a file, called a trajectory file, from which one can deduce a significant amount of information regarding the investigated system.

3.4 Broadband Dielectric Spectroscopy

Broadband dielectric spectroscopy (BDS) is a technique in which a sample is placed within an electric field oscillating at relatively low frequencies ($\sim 10^{-2}$ – 10^{12} Hz), thus probing relatively slow atomic and molecular motions, such as dipole reorientations or charge transport. Due to its broad range of low frequencies it is also excellent at studying e.g. molecular dynamics below the glass transition temperature, and to characterize the dynamics which determine the glass transition itself.

In a BDS experiment, the sample is typically placed between two electrodes, effectively producing a capacitor (as seen in figure 3.4), with the sample acting as the dielectric material of that capacitor. By measuring the responding elec-

trical field from an applied electric field of this equivalent circuit, it is possible to measure the dielectric permittivity of the material as a function of the frequency of the applied field. Based on these dielectric properties of the material it is possible to draw conclusions about the molecular properties of the sample.

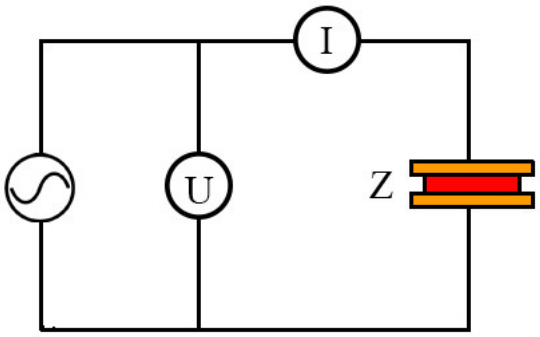


Figure 3.4: Simplified electric circuit of a BDS setup. The sample (red) is placed as the dielectric material in between the electrodes (yellow). By applying a known voltage (U) and measuring the current (I) the complex impedance (Z) is indirectly measured.

This section provides a brief overview of dielectrics in electric fields, how certain molecular dynamics effect the dielectric properties of the material, and how this is taken advantage of in the BDS method.*

3.4.1 Theory

When an electric field is applied to a sample, the charge density of the atoms within that sample will shift in accordance with the electric field. The field created by the material is typically called the polarization field ($\mathbf{P}$), and will be proportional to $\mathbf{E}$ such that:† $\mathbf{P} = \chi_s \varepsilon_0 \mathbf{E}$, where χ_s is the static electric susceptibility and ϵ_0 is the electric permittivity of vacuum ($8.854{\cdot}10^{-12}\frac{[As]}{[Vm]}$). The dielectric displacement vector $\mathbf{D}$ is the electric field after corrections by

*The theory presented in this section is based on a combination of different sources, most notably Refs. 125 and 126

†Given that the material of the sample is isotropic and homogenous

the polarization, i.e.:

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} = \varepsilon_0 \varepsilon_s \mathbf{E} \tag{3.33}$$

Where ε_s is the static dielectric permittivity of the sample defined as $1 + \chi_s$. The latter part of equation 3.33 holds since $\mathbf{E}$ and $\mathbf{P}$ both point in the same direction. ε_s is a material property, and is a measure of how an electric field changes due to the presence of the material. Within a material under an electric field, atoms and dipoles will tend to align in accordance to the electric field, under multiple different circumstances, such as transport of charges (conductivity) or reorientation of permanent (or induced) dipoles.

When the electric field is time dependent, such that $E^*(t) = E_0 e^{i(\omega t)}$ (where ω is the radial frequency, and E_0 is the amplitude of the field), it is important to take into account that the electric permittivity will no longer be static but also depend on the oscillation of the field.[125] Thus, describing the dielectric displacement field as a function of frequency rather than time:

$$\mathbf{D}(\omega) = \varepsilon_0 \varepsilon^*(\omega) \mathbf{E}(\omega) \tag{3.34}$$

where $\varepsilon^*(\omega)$ is the complex dielectric permittivity of the sample with a real (ε') and imaginary (ε'') component ($\varepsilon^*(\omega) = \varepsilon'(\omega) - i\varepsilon''(\omega)$). By measuring a spectra of $\varepsilon^*(\omega)$ over a range of frequencies it is possible to deduce several different material properties of your sample (as will be discussed more below). But first, let us briefly mention how $\varepsilon^*(\omega)$ is obtained. From the simplified circuit shown in figure 3.4, it can be seen that the complex impedance of the sample (Z^*) can be obtained via Ohm's law ($Z^* = \frac{U^*}{I^*}$), which, assuming the impedance is purely due to capacitance, relates to the complex capacitance (C) as: $Z^* = -\frac{i}{\omega C}$. When the capacitor does not contain any dielectric material, its capacitance is fairly straight-forward to calculate by $C_0 = \frac{A\varepsilon_0}{d}$, where A is the area of the capacitive plates, and d is the distance between the plates. With the dielectric material in, the capacitance becomes $C = \varepsilon^*(\omega) C_0$, where $\varepsilon^*(\omega)$

is the complex frequency dependent permittivity, which we then can relate to the measured impedance by the relationship:

$$\varepsilon^*(\omega) = -\frac{i}{Z^*(\omega)\omega C_0} \tag{3.35}$$

However not all materials exhibit perfect insulating capacitance behaviour; charge-transfer could also occur inside the sample, which gives rise to a conductivity term to the total impedance, such that the total measured permittivity (ε_{tot}) of the sample can be expressed as:

$$\varepsilon_{tot}(\omega) = \varepsilon^*(\omega) - i\frac{\sigma_0}{\varepsilon_0\omega} \tag{3.36}$$

where σ_0 is the DC-conductivity.

3.4.2 Material response functions

In the section above, it was explained how the measured complex electric impedance can be related to the complex permittivity of the sample. Here it will be shown how this quantity relates to certain properties of the material, such as dipole relaxations.*

As previously mentioned, when an electric field is applied to a material of dipoles (permanent or induced), these dipoles will tend to align to the electric field, creating a polarized field, related to the complex electric permittivity as discussed above. If the electric field is suddenly switched off, the dipoles will not respond instantaneously, but rather relax over time to their original random orientations due to thermal fluctuations. The time of this relaxation (τ) depends on the local environment of the individual dipoles. Applying an oscillating field varying frequencies, the frequency dependence of the complex permittivity becomes:

*Virtually all material properties could in principle be encoded in ε^*, but here we will focus on those accessible for BDS measurements.

$$\varepsilon^*(\omega) = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + i\omega\tau} \tag{3.37}$$

Where ε_∞ is the permittivity when the frequency goes to infinity. The real (ε') and imaginary (ε'') parts are:

$$\varepsilon'(\omega) = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + (\omega\tau)^2} \tag{3.38}$$

and

$$\varepsilon''(\omega) = \frac{(\varepsilon_s - \varepsilon_\infty)\omega\tau}{1 + (\omega\tau)^2} \tag{3.39}$$

Equation 3.37 is called the Debye equation, and it expresses the characteristic relaxation time (τ), and amplitude ($\Delta\varepsilon = \varepsilon_s - \varepsilon_\infty$) of a specific dipole relaxation, which can be measured by e.g. a BDS experiment. In most amorphous materials however, molecular dipoles seldom have exactly the same characteristic relaxation time; some dipoles may be slightly more or less bonded to a co-solute than others for example. Instead of a Debye model a dipole relaxation can often be advantageously described by multiple relaxations, statistically distributed around a certain average relaxation time. This is typically done by either a symmetric broadening of the Debye equation, using the Cole-Cole equation, or an asymmetric broadening of the Debye-function, using the Cole-Davidsson equation. Mathematically, these two can both be written using the so called Havriliak Negami (HN) function[127]:

$$\varepsilon^*(\omega) = \varepsilon_\infty + \frac{\Delta\varepsilon}{(1 + (i\omega\tau)^\alpha)^\beta} \tag{3.40}$$

α, and β are shape-parameters, where if β is set to 1, but α varies between 0 and 1, one obtains a symmetrically broadened Cole-Cole equation, if α is set to 1, but β varies between 0 and 1 one obtains the asymmetrically broadened Cole-Davidsson equation, and if both α and β are set to 1, one obtains the Debye equation again (illustrated in figure 3.5).

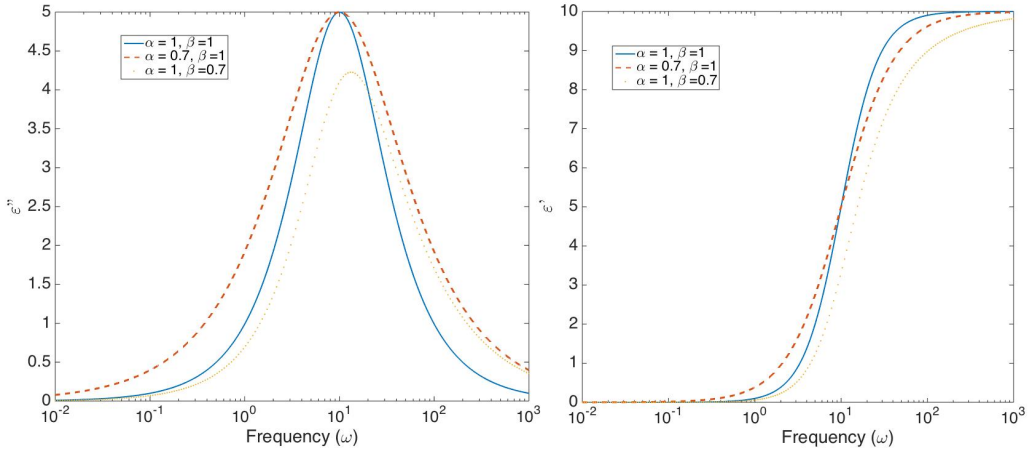


Figure 3.5: Real (right) and imaginary (left) parts of the complex permittivity function for different HN-functions. Blue lines represent a Debye function where both α and β are 1, red dashed lines represent a symmetrically broadened Cole-Cole equation, and the black dotted lines represent an asymmetrically broadened Cole-Davidsson equation

3.4.3 Temperature dependence

For an isothermal BDS experiment, it is common to detect a couple of relaxation functions (and conductivity for higher temperatures, where ion diffusion is possible) within the experimental frequency window. Based on the framework presented above, the permittivity of the sample could then in theory be fitted with a sum of relaxation functions, as described by the Havriliak Negami (HN) function (equation 3.40), plus the conductivity term (as from equation 3.36):

$$\varepsilon_{tot}(\omega) = i\frac{\sigma_0}{\varepsilon_0\omega} + \sum_k \left(\frac{\Delta\varepsilon_k}{(1+(i\omega\tau_k)^{\alpha_k})^{\beta_k}} + \varepsilon_{k\infty} \right) \tag{3.41}$$

The resulting fit of an isothermal measurement is illustrated in figure 3.6.

By fitting all relaxation functions for a series of temperatures, it is possible to deduce certain important properties of the different relaxation functions. In figure 3.7, τ for a couple of relaxations in a sample is plotted as a function of $\frac{1000}{Temperature}$. In figure 3.7 the process labelled "Dipole relaxation" is linear,

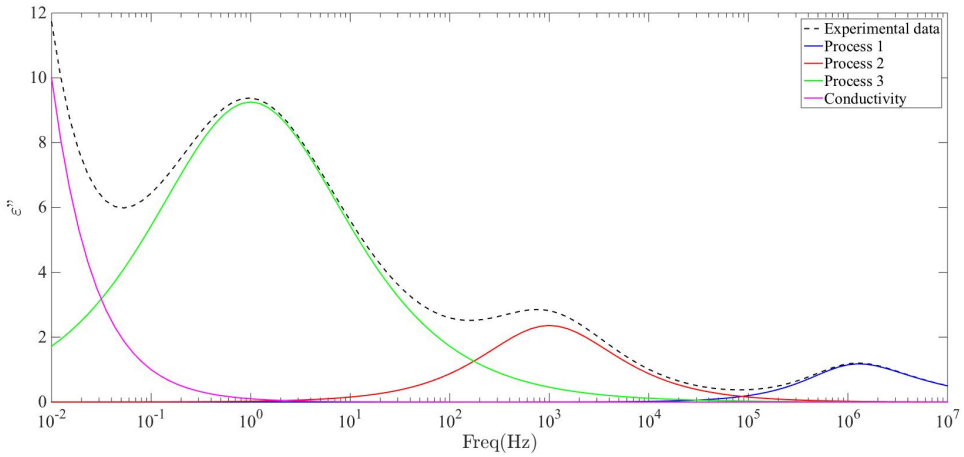


Figure 3.6: Fit of a fictional BDS measurement (black dashed curve) with two Cole-Cole-processes (green and red curve), one Cole-Davidsson process (blue curve) and one conductivity term (magenta).

and is called an Arrhenius process*. An Arrhenius process describes a simple relationship between the relaxation time of a molecular process with its temperature, namely:

$$\tau = \tau_0 exp(-\frac{E_a}{k_b T}) \tag{3.42}$$

Where k_b is Boltzmann's constant, τ_0 is the relaxation time when extrapolating the temperature (T) to infinity, and E_a is the activation energy of the relaxation process. By fitting an Arrhenius process, it is thus possible of obtaining information about the general rate of the relaxation through τ_0, and about the bond-energy which is required to be broken in order to e.g. rotate a molecule (through E_a).

The two curved slopes in figure 3.7 labelled as "α-relaxations" are often referred to as non-Arrhenius-, or α-processes. These processes are typically cooperative molecular motions, related to the viscosity and the glass transition of the material. When decreasing the temperature, these processes deviate towards infinite relaxation times, and at the calorimetric T_g, the relaxation time of these processes are around 100 s. These can be fitted using the so

*This type of figure are actually typically referred to as Arrhenius plots

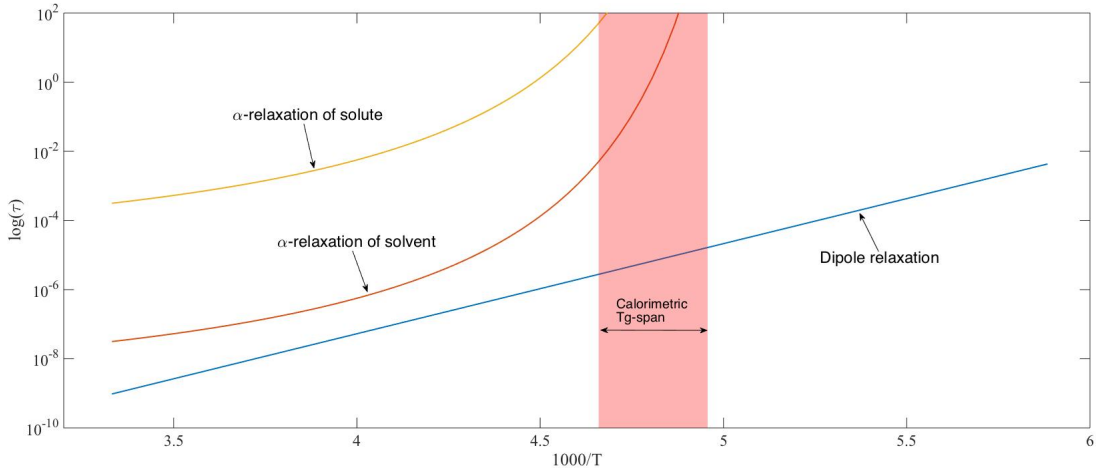


Figure 3.7: Arrhenius plot of three different relaxations of a sample. Blue curve is an Arrhenius process coming from a dipole relaxation. Yellow and red curves are VFT curves originating from collective dynamics of solvent or solute molecules respectively.

called "Vogel-Fulcher-Tammann" (VFT) equation:

$$\tau = \tau_0 exp(\frac{DT_0}{T - T_0}) \tag{3.43}$$

Where T_0 is the temperature where the relaxation time diverges towards infinity, and D is the fragility parameter. D is a parameter often used to quantify how "strong" a glass is; if it is low, the process deviates more from Arrhenius-behaviour, and the glass that forms is classified as "fragile", and if it is more Arrhenius-like it is classified as "strong".

3.5 Differential Scanning Calorimetry

Differential Scanning Calorimetry (DSC) is a method for determining different types of thermal events in a material. The principle behind this method is to change the temperature of a sample and measuring the involved heat (enthalpy). This process enables the experimenter to detect e.g. glass transitions, crystallization, denaturation, etc.

Inside the DSC cell there are two sample platforms; one for the sample and one for a reference sample. The reference sample is typically an empty sample holder, identical to the one holding the sample under investigation. When the temperature is set to change a certain rate ΔT, the sensors on the sample and on the reference sample registers the difference in heat flow to and from the two samples ΔQ. The difference ΔQ corresponds to the amount of energy required to change the temperature of the sample by ΔT. Another important investigated property is the specific heat capacity, obtained after normalisation with respect to e.g. sample mass and baseline. The heat flow to and from a sample is typically constant when no physical or chemical changes occur. However, when such events do occur, they show up in the signal in different ways depending on the nature of the event. In Figure 3.8 some events are pointed out that are relevant for the present thesis and will be briefly discussed here.

1. **Crystallization**. When the sample crystallises it transforms from a highly disordered state to a more ordered one, the entropy (S) decreases and thus the process is exothermic, which shows up as a positive peak as indicated in Figure 3.8.

2. **Melting**. Opposite of crystallization, the disorder of the sample is increased more energy is required to break the crystal structure, which shows up as an endothermic dip in the DSC curve.

3. **Glass transition**. The glass transition, is a second-order process (ex-

plained in section 2.1), and it is thus associated with a change in the heat capacity of the sample. It can be seen as a step in the baseline of the heat flow.

4. **Denaturation**. During the denaturation process the hydrogen bonds keeping the protein in its functional state are broken. When this happens the protein can easily unfold and aggregate. This is an irreversible process and shows up as an endothermal peak in the DSC scan.

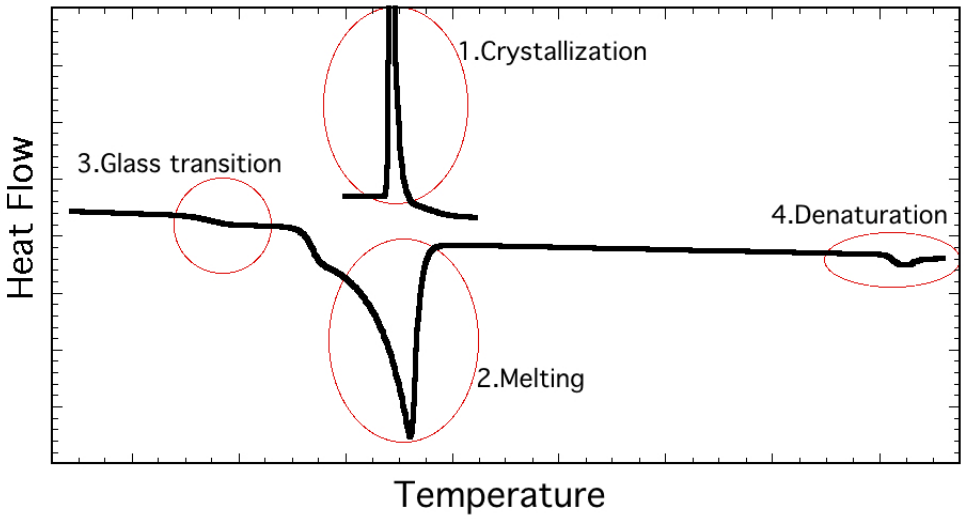


Figure 3.8: Typical DSC scan of a sample containing protein, trehalose, and water. (1) Sample partially crystallise during cooling. (2) During heating, the crystalline part of the sample melts. (3) A part of the sample exhibits a glass transition. (4) The protein denatures.

4

Experimental Procedures

4.1 Neutron Diffraction Experiments

4.1.1 Diffractometer

The diffraction data presented in papers I, II, and IV were obtained using the NIMROD (Near and InterMediate Range Order Diffractometer) diffractometer[128] at the spallation source ISIS, Rutherford Appleton Laboratory, UK. The incident neutrons at NIMROD arrive to the sample in wavelengths between 0.05–10 Å, and it has a broad Q-range of 0.01–50 $Å^{-1}$. This broad Q-range makes NIMROD specialized in measuring a very wide range of length-scales, from less than 1 Å, up to more than 300 Å. It is thus possible to combine structural data from short-range distances (through the large scattering angles) to longer range distances (via the small scattering angles). This makes NIMROD suitable for probing large-scale structures inside a disordered medium, such as liquids or macromolecules in solutions.

4.1.2 Data Corrections

To obtain useful structure factors from the raw neutron diffraction data, the GUDRUN data correction software has been used in the work presented in this

thesis. More details about the discussed corrections can be found in Ref. 11.

Neutron diffraction data is typically normalised using a vanadium plate since it exhibits almost only incoherent scattering. Therefore, the scattering from the vanadium plate does not depend on Q, and thus show up as a uniform background signal for a given incident neutron flux. This background signal is used to give the data an absolute scale, since the theoretical scattering of the vanadium plate is relatively easy to calculate. The number of detected neutrons at each detector can thus be divided by the number of theoretically detected neutrons to give the normalisation factor for that particular detector.

Measurements are normally also made on an empty sample holder, in order to subtract the signal from the sample holder from the measured data.

- **Multiple scattering corrections.**

  In the theory presented about neutron scattering above, it is assumed that a scattered neutron travels directly from the sample to the detector. This is however merely an approximation; multiple scattering events are possible, and sometimes this has to be taken into account. The probability that a neutron is scattered multiple times can be calculated from the atomic composition of the sample and the sample geometry. From these parameters, a background signal can be calculated and subsequently removed from the final structure factor.

- **Absorption corrections.** Another approximation in the theory presented in section 3.1.1 is that there is no absorption of neutrons in the sample. The cross section of a sample was presented as a *scattering* cross section. However in reality, the total cross section should include the absorption cross section ($\sigma_{tot} = \sigma_{scattering} + \sigma_{absorption}$). For the energies of the incident neutrons in the present experiments, the absorption cross section is assumed to have a linear dependency to the wavelength of the incident neutrons, which is a good approximation.[11] Similar to the

situation with the multiple scattering corrections, the absorption is estimated from the atomic composition of the sample plus sample holder, and their combined geometry. The atoms are furthermore approximated to be isotropically distributed within the sample (as well as in the sample holder).[11]

- **Deadtime corrections.**

  Deadtime corrections are necessary due to the "deadtime" of the detectors. When a detector is hit by a neutron, it requires some time before it is able to count the next one. Thus, they may count multiple hits as just one if they are unable to resolve these in time. GUDRUN takes this into account and compensates for such effects.

4.1.3 Isotope substitution

As previously discussed in section 3.1, the structure factor, $S(Q)$, can be expressed as a linear combination of the different partial structure factors, $S_{\alpha\beta}(Q)$, as in equation 3.14. In theory, this equation can be inverted, with the help of substitution of isotopes. Given that a molecular system has the same structure for different isotopes, the only thing differing in the obtained structure factors is the scattering lengths. By choosing appropriate substitutions different partial structure factors can therefore be emphasized or hidden, thus in total yielding a full set of partial structure factors (which subsequently generates the partial pair correlation functions through Fourier transformation). The number of substitutions required to fully solve this in a sample of n different chemical species is the same as the number of different partial structure factors that can be generated, i.e. $\frac{n(n+1)}{2}$.

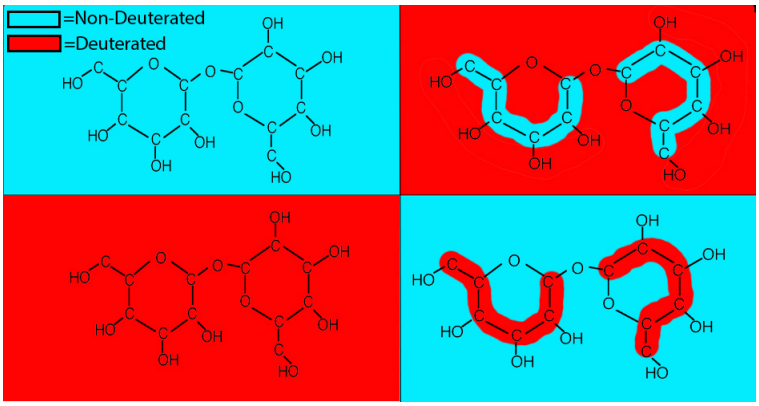


Figure 4.1: Sketch showing how isotope substitution gives contrast to particular sets of atoms in a sample.

4.2 EPSR Modeling

4.2.1 Coordination numbers

The coordination number, $n_\beta^\alpha(r_1, r_2)$, which gives the number of atoms of type β surrounding a central atom type α within a radial distance between r_1 and r_2, is calculated using the partial pair correlation functions $g_{\alpha\beta}(r)$:

$$n_\beta^\alpha(r_1, r_2) = 4\pi c_\beta \int_{r_1}^{r_2} g_{\alpha\beta}(r) r^2 dr \tag{4.1}$$

where c_β is the atomic number density of atom β, and $g_{\alpha\beta}(r)$ is defined through equation 2.4.

This number is in general reported for different coordination shells (first, second, etc.), where e.g. the first coordination shell is typically defined as the coordination number between $r_1 = 0$ and r_2. r_2 is the distance where the first minima appear in $g_{\alpha\beta}(r)$ after the first obvious peak (in the O-O correlation for water in Figure 2.4 for example, r_2 would be approximately 3.4 Å).

4.2.2 Clusters

In order to calculate the cluster size distribution in the molecular model, the cluster-subroutine in EPSR was used. This subroutine counts the number of molecules in each cluster, ranging from a non-clustered molecule with a cluster size of 1, to a fully clustered system, where all specified molecules in the model are connected to the same cluster. In papers I, II and IV, the definition for two trehalose molecules to be considered clustered was that the minimum distance between an oxygen and a hydrogen from different trehalose molecules was 2.5 Å. If yet another trehalose bind to one of the already clustered trehalose molecules, the cluster-size for this cluster grow with one unit.

4.2.3 Angle distributions

Another important EPSR subroutine used in this work is one that counts various angle distributions. Within this subroutine the user defines triplets of atoms, and maximum distance criteria for these atoms. The program then measures the angle between every triplet of atoms that satisfies the distance criteria. This subroutine was used to investigate the geometric configuration of the water structure in papers I and II (see Figure 4 and 5 in paper I and Figure 4 in paper II).

4.2.4 Hydrogen bonding

The number of intermolecular hydrogen bonds between water and trehalose were calculated for the structures obtained from the EPSR simulation. The calculation was done using different sets of criteria for different definitions of a hydrogen bond. Previous EPSR studies (such as Refs. 129 and 130) in the literature have used a maximum distance criterion, i.e. a hydrogen bond is defined by that the distance between a donor hydrogen and an acceptor oxygen is less than 2.5 Å. However various MD-studies (e.g. Refs. 39 or 131)

typically employ a distance criteria combined with an angular criterion. The most common setup for these two criteria is that the minimum distance between an acceptor oxygen and a donor oxygen is less than 3.4 Å and that the O - H - - - O angle is maximum 120°. Other criteria were also used in papers I and II in order to compare with previous studies.
In papers I and II, the hydrogen bond calculations were done for a large set of different configurations (∼1000 configurations) of the system after the system exhibited the best fit to the experimental data. In paper I this was done with the use of a homemade Matlab script, and in paper II all configurations were converted into a pseudo-trajectory-file which could be analyzed with the help of standard MD-analysis software, such as VMD[132] or Gromacs.

4.2.5 EPSR Modeling of proteins

The aforementioned pseudo-trajectory-files were also created for the analysis of the three-component systems presented in paper IV. This was mainly analyzed using VMD, which simplified the calculation of e.g. partial correlation functions of specific selections of atoms, such as protein-surface atoms to water oxygen correlations.

4.3 QENS procedure and data analysis

4.3.1 Measurement

All QENS measurements were done on the time-of-flight spectrometer IRIS (ISIS, Rutherford Appleton Laboratory, UK)[133]. The principle of a time-of-flight spectrometer is (as the name implies) to measure the variation in the time of arrival at the detectors. For elastic scattering the arrival time is known, however if the neutron gain or lose energy during scattering, this time will differ

and can be related to the energy-difference as:[134]

$$\Delta E = E_1 - E_2 = \frac{1}{2} m_n \left[\left(\frac{L_1}{t - t_2}^2 - \frac{L_2}{t_2}^2 \right) \right] \tag{4.2}$$

Where m_n is the neutron mass, L_1 is the known path-length between the neutron source and the sample, L_2 is the path-length between the sample and the detector, t is the total time-of-flight, and t_2 is the travel-time of a neutron from the sample to the detector.

The samples were placed in annular aluminium alloy cans with 0.1 or 0.25 mm sample thickness; and the scattering contribution from these cans were subtracted before data analysis.

4.3.2 Fitting procedure

Although basic theory about fitting QENS data was presented in section 3.1.2, the fitting procedure used in paper V was a slightly modified version of this procedure.

Instead of fitting each of the three isotope compositions (for three-component sample) separately with a sum of stretched exponentials (as in equation 3.22), the data from all three isotope compositions (denoted as $I^1(Q,t)$ $I^2(Q,t)$ and $I^3(Q,t)$)were fitted simultaneously by solving the following non-linear equation system:*

$$\begin{bmatrix} A_P^1 & A_T^1 & A_W^1 \\ A_P^2 & A_T^2 & A_W^2 \\ A_P^3 & A_T^3 & A_W^3 \end{bmatrix} \cdot \begin{bmatrix} e^{-(t/\tau_P)^{\beta_P}} \\ e^{-(t/\tau_T)^{\beta_T}} \\ e^{-(t/\tau_W)^{\beta_W}} \end{bmatrix} = \begin{bmatrix} I^1(Q,t) \\ I^2(Q,t) \\ I^3(Q,t) \end{bmatrix} \tag{4.3}$$

Where the subscripts P, T, and W stand for protein, trehalose, and water and represent which molecular component each number refers to. The superscripts signifies which of the three isotope compositions (labeled 1, 2 and 3) it represents. The A values refers to the relative scattering contribution of that

*The solution was obtained using the built-in "lsqcurvefit"-function in MATLAB, which implements a rrust-region-reflective least squares algorithm.

component within its particular isotope composition, and are assumed to be known values (see Table 1 in paper V). τ and β are the relaxation time and stretching parameter respectively of a specific molecular component, which are assumed to be independent on isotope composition.

4.4 BDS setup and fitting procedure

All BDS measurements were done using an Alpha-S High Resolution Dielectric Analyzer (from Novocontrol) covering a frequency range of 10^{-3}–10^{7}Hz. The samples were placed, together with a small distance-spacer of teflon or silica (with a typical thickness of 100μm), between two gold-plated electrodes with a diameter of 20mm.

The samples were quenched within the sample chamber of the spectrometer with a stream of cold nitrogen gas (∼120K, with a cooling rate of approximately 20K/min) and then, for the isothermal measurements, heated with 5-10K between each frequency measurement sweep, up to about 300K. In certain cases, where some relaxations were unclear, isochronal measurements were also performed in which the samples were quenched to 120K and then heated continuously with 0.2K/minute, and the response were measured for eight different frequency-points between 10^{-1} and 10^{7}Hz.

4.4.1 Fitting

In order to fit the dielectric data several approaches were used. The first step was usually to determine which relaxation processes were present. This was done at first by simply observing the data for different data representations, such as $\varepsilon'(\omega)$, $\varepsilon''(\omega)$, or $\tan(\delta)$ (defined as $\tan(\delta)= \frac{\varepsilon''(\omega)}{\varepsilon'(\omega)}$. The $\tan(\delta)$-representation usually reveal relaxations which for example in $\varepsilon''(\omega)$ could be hidden by conductivity, however the actual fitting procedure of $\tan(\delta)$ requires more parameters and thus becomes tricky to fit directly. Another common

method of detecting relaxation processes which are difficult to detect is to study the Ohmic-conduction-free dielectric loss ε''_{rel}, which can approximately be obtained by taking the logarithmic derivative of the real part of the dielectric permittivity[135]:

$$\varepsilon''_{rel} \approx -\frac{\pi}{2}\frac{\delta\varepsilon'(\omega)}{\delta ln\omega} \tag{4.4}$$

Additionally, it has been shown that additional isochronal measurements are more sensitive to weak relaxations than isothermal measurements.[136,137] For this purpose, this measurements was occasionally made to ensure the existence of certain unclear processes.

Once approximate frequencies for the different relaxation processes at different temperatures were determined, the fitting procedure continued by fitting each process with a HN-function* (mainly by the use of the Novocontrol's fitting software WinFit[138]).

4.5 DSC Experiments

All DSC experiments for this thesis were performed on a DSC Q1000 (TA Instruments). For the measurements, each of the samples were placed in hermetically sealed aluminum pans, which were placed on the sample podium within the DSC-sample chamber. The DSC-sample chamber is coupled to a liquid nitrogen cooling system, which allows for cooling to temperatures of -180°C. The DSC-sample chamber is also connected to a thermal element, which could heat the samples up to 550°C. The temperature and the heat flow to and from the reference and the sample are measured independently.

Before the measurements, heat capacity calibration was performed using two sapphire disks (one for the reference podium and one for the sample podium). By scanning the heat flow and the heating rate for a large span of temperatures it is possible to find the calibration constant, K, defined as

*Typically a Cole-Cole, or a Cole-Davidsson function are used, to be more precise.

$C_p = K\frac{Heatflow}{HeatingRate}$,[139] since the heat capacity of sapphire (C_p) is well-known and stable. A temperature and enthalpy calibration was also performed by heating three standard substances (indium, water, and mercury) over their melting temperatures. The melting temperatures and heat of fusion for these substances are well known, and thus any deviations in temperature or enthalpy measurements can be corrected for by using these calibration constants.

5

Summary of Appended Papers

Paper I: Structure of aqueous trehalose solution by neutron diffraction and structural modeling

In paper I, some of the fundamental properties of the structure of aqueous trehalose were investigated. Many of the previous reported studies on these properties have been done by e.g. MD simulations (see for example Ref. 39 and 43). Such studies are sometimes limited by the approximative force fields, which do not necessarily produce correct structures. For the work in this study, the structure was directly investigated by the use of neutron diffraction. The data were collected on the NIMROD neutron diffractometer, and were thereafter analyzed using EPSR simulations. A similar study has previously been performed,[51] however with a different isotope substitution scheme than the one performed in this study. In Ref. 51, the authors deuterated the exchangeable trehalose hydrogens, which thereafter were dissolved in H_2O. However, the presence of H_2O makes the deuterated hydrogen groups exchange back to hydrogen, thus resulting in just varying degrees of deuteration of the exchangeable hydrogens. In paper I, the exchangeable hydrogens of trehalose was first exchanged to either be fully hydrogenated, fully deuterated, or with half hy-

drogenated half deuterated, and subsequently dissolved in H_2O, D_2O, or HDO respectively. Furthermore, in three (out of six) of the samples we used trehalose which had their non-exchangeable hydrogen-groups deuterated. These deuterated trehalose samples were also dissolved in either H_2O, D_2O, or HDO, with their exchangeable hydrogens substituted to match their respective aqueous environment.

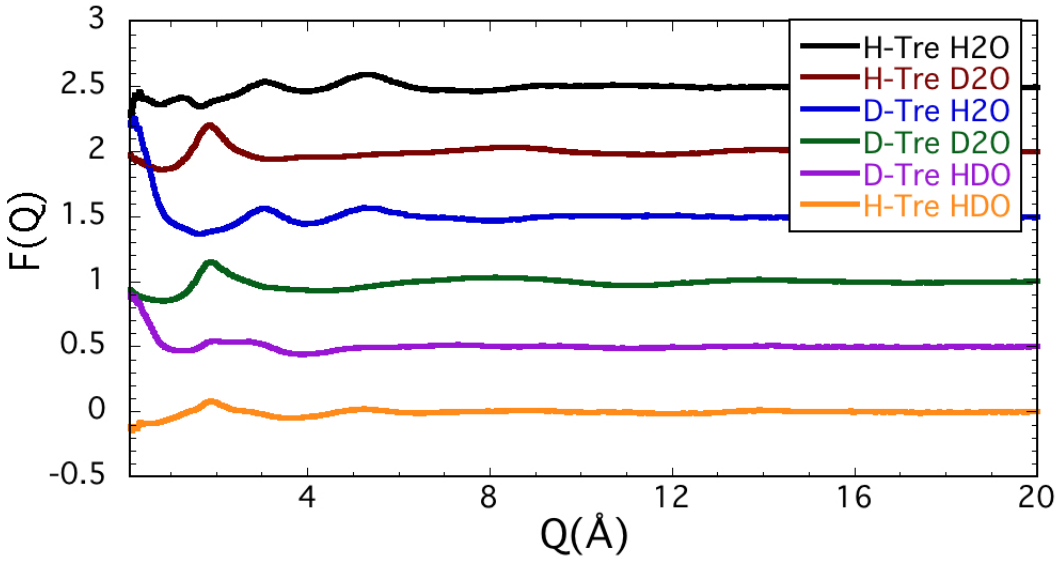


Figure 5.1: Neutron diffraction data of different isotope compositions of trehalose in water

The EPSR modeling was performed by building a simulation box of 2000 water molecules and 52 trehalose molecules. Analysis of the structure of water in the presence of trehalose indicated that the water structure is significantly perturbed, most likely due to that trehalose exposes a large portion of its potential hydrogen-bonding sites for the water molecules to bind to. This finding presumably also correlates with the low degree of clustering of trehalose molecules that was found, which may play a large role in explaining the peculiarity of trehalose.

Paper II: Structural comparison between sucrose and trehalose in aqueous solution

Paper II is a continuation on the work published in paper I, in which the previously obtained trehalose neutron diffraction data is compared to neutron diffraction data of aqueous sucrose samples, with the same water content and isotope composition as in paper I. These two rather similar disaccharides exhibit certain quite different properties, which is hypothesized to be important to understand why trehalose display better stabilizing properties when preserving biological materials. The two sets of data showed however surprisingly small differences. These data-sets are compared to that of bulk water neutron diffraction data, and the differences between the two disaccharides was, based on this comparison, hypothesized to be due to a larger perturbation of the water structure in the case of trehalose.

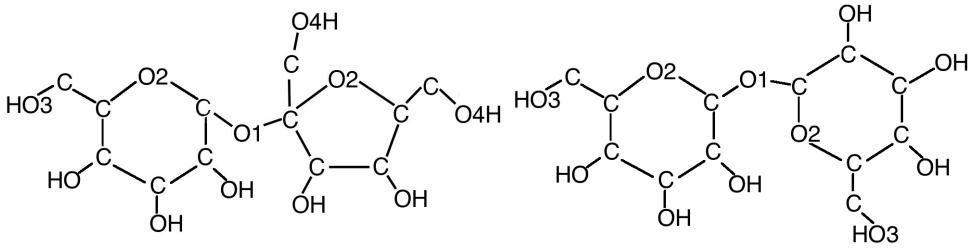


Figure 5.2: Molecular structure of sucrose (left) and trehalose (right)

The sucrose data was also modeled using EPSR much like in paper I. A twice as large simulation box compared to paper I was created however, alongside with a new revised EPSR model of trehalose, such that both models were created with an equal setup. From these models the analysis of the diffraction data could be supported: the two disaccharides exhibit similar water/disaccharide interactions, but the water structure in trehalose is perturbed more than in the sucrose sample (although slightly less than reported in paper I). Apart from this we obtained similar results regarding e.g. hydrogen bonds and clustering as for trehalose in paper I.

Paper III: The Role of Trehalose for the Stabilization of Proteins

In paper III, the role of trehalose during the stabilization process was investigated by differential scanning calorimetry. The samples studied were composed of mixtures of water, trehalose, and myoglobin, at a wide range of concentrations, typically at a water concentration less than 75wt%. Particularly, the glass transition temperatures and the denaturation temperatures were studied to correlate the stability of the glassy matrix (as determined by T_g) with the stability of the protein (determined by T_{den}). It was shown that more trehalose was correlated with both an increment in T_g and in T_{den}.

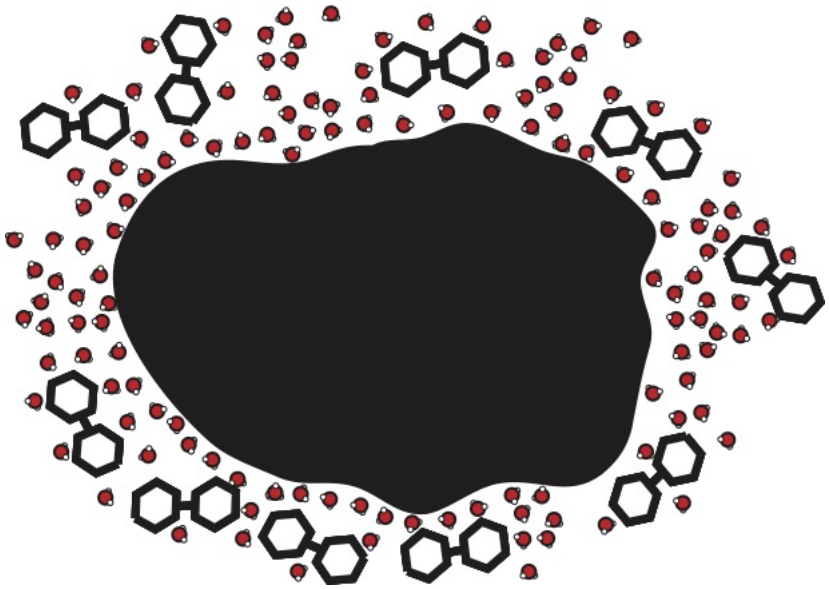


Figure 5.3: Sketch of preferential hydration interpretation

Samples which did not form ice exhibited a positive correlation between T_g and an increase in protein concentration, but a negative correlation if the sample did form ice. This is explained by that the protein effectively dries out the glassy trehalose–water matrix when there is no ice formation, which thus raises T_g. If ice is formed however, the amorphous part of the sample becomes freeze-concentrated by expelling any excess water molecules. In this

freeze-concentrated part, T_g is determined by the protein:trehalose ratio, and the protein has a lower T_g than trehalose.

By determining the maximum water concentration before the sample exhibit crystallization, for different trehalose:protein ratios, it was found that water preferentially adsorbs to the protein surface. Thus, the results indicate that the preferential hydration model (as sketched in Figure 5.3 and explained in section 2.5.3) is more likely than the water replacement model, for the investigated concentrations.

Paper IV: Structural Role of Trehalose for Protein Stabilization and Inhibiting Protein Aggregation from Neutron Diffraction

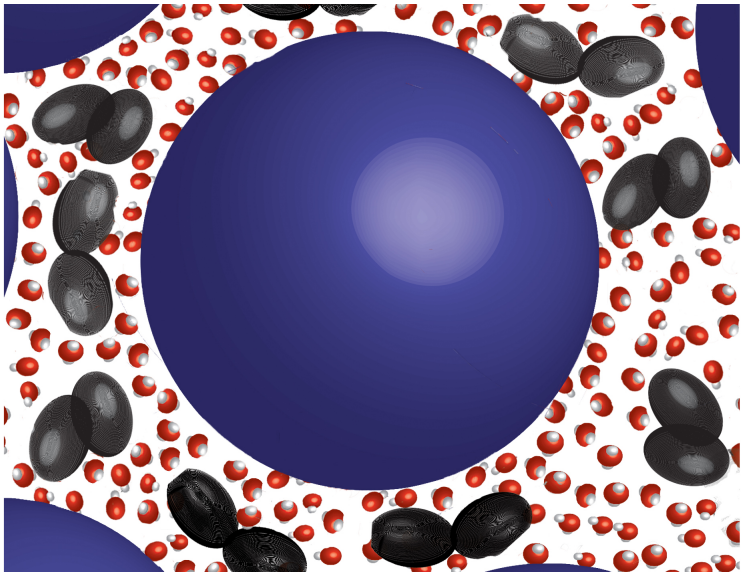


Figure 5.4: Sketch of proteins (blue) preferentially hydrated with water molecules. Trehalose molecules (black) are excluded from the protein surface.

The preferential hydration hypothesis was further tested in paper IV, however in a more diluted system of water, trehalose and myoglobin (50wt% water, and 25wt% of both trehalose and myoglobin, based on a fully hydrogenated

sample), than in paper III. In this paper the structure of this three-component system was elucidated using neutron diffraction measurements on the NIMROD diffractometer, and analyzed with the use of EPSR modeling. Six different isotope compositions of this three-component system was made, and three isotope compositions of a two-composition system containing myoglobin in water (66wt% water, and 33wt% myoglobin, based on a fully hydrogenated sample). By analyzing the small-angle scattering of both the three- and two-component systems, it was concluded that the presence of trehalose had an inhibiting effect on protein aggregation. This effect was coupled to the importance of trehalose as a treatment against neurodegenerative diseases, such as Alzheimer's and Huntington's disease, as reported in e.g. Refs. 102 and 103. The wide-angle scattering part was fitted using EPSR modeling. In this model a single protein in a water/trehalose solution (or just in water for the binary water/protein case), corresponding to the same concentration as the samples, was simulated and fitted to the experimental data. From the fitted model the structure of water and trehalose around the protein could be quantified. This model was compared to a hard-sphere model, where the water and trehalose molecules were randomly oriented around the protein. It was concluded that the diffraction data was best fitted with a model in which the trehalose molecules were preferentially excluded from interacting directly with the protein surface. This result indicates that the protein is mainly not stabilized by a direct trehalose-protein coupling, but rather via indirect interactions through the hydration layer, supporting the theories surrounding the preferential hydration model as discussed in section 2.5.3.

Paper V: Mechanism of Trehalose Induced Protein Stabilization from Quasielastic Neutron Scattering and Molecular Dynamics Simulations

In paper V, the dynamical properties of the same systems as studied in paper I and IV were investigated using QENS and MD simulations. QENS measurements on three of the isotope compositions used in paper IV for the three-component system, and two isotope composition from either two-component system, were performed using the IRIS spectrometer. These data were converted into intermediate scattering functions and fitted with a sum of Kohlrausch-Williams-Watts (KWW) stretched exponential functions.

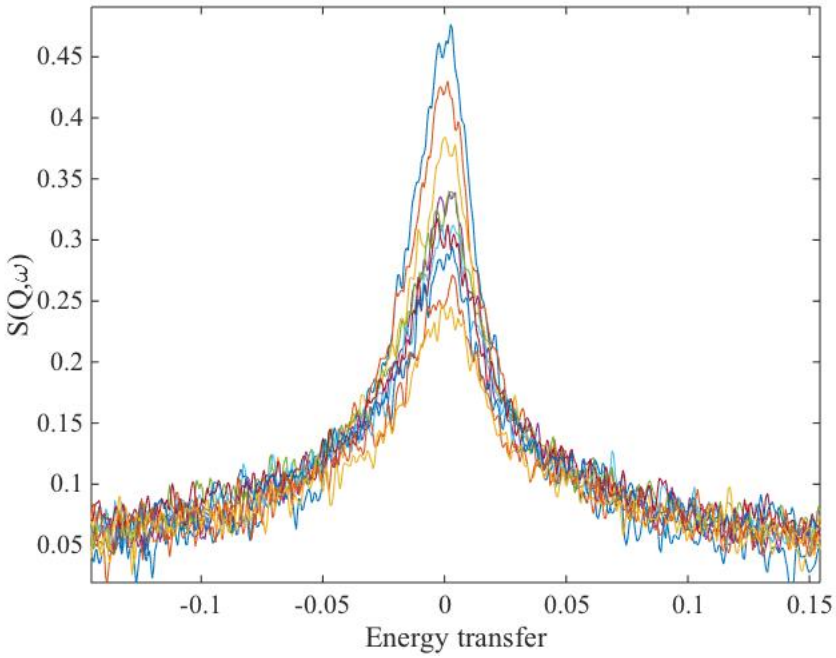


Figure 5.5: QENS data of myoglobin in H_2O and trehalose.

Each molecule type was assigned an individual KWW-function, and the amplitude of this process was set to the relative incoherent scattering contribution. All isotope compositions for a sample were fitted simultaneously with the same relaxation times and stretch parameters. From the resulting fits, the

local diffusion time of the individual molecules in the samples were calculated, and it was concluded that the local diffusion of both the water and the protein was retarded to a greater extent in the presence of trehalose. The diffusion of trehalose however, was relatively unchanged in the three-component system compared to the two-component system. This suggests that the trehalose is not directly affected by the slower protein dynamics, and thus does do not directly bind to the protein surface.

From the MD simulations, the relaxation times of individual molecules were calculated. Using this method it was also found that the protein preferentially excluded the trehalose molecules. Furthermore, in agreement with the QENS results, the MD analysis showed that the presence of trehalose reduced the dynamics of the protein residues, and that the water dynamics was retarded by both the presence of trehalose and protein, but to a higher degree in the three-component system. However, the MD data showed that trehalose did slow down in the three-component system compared to in the two-component system. This discrepancy between QENS and MD data is hypothesized to be either due to the MD simulation yielding more trehalose-protein interactions than what is observed experimentally, or that the presence of these protein-bonded trehalose molecules are too few and/or too slow to be seen in the QENS data.

Paper VI: Dielectric spectroscopy study of proteins embedded in trehalose and water

Dynamical properties of myoglobin in relatively dry systems of water and trehalose were investigated in paper VI, by the use of broadband dielectric spectroscopy. The purpose of this study was to elucidate how trehalose altered different relaxation processes of the solvent and the myoglobin. Protein stabilization can be achieved by slowing down and decreasing the amplitude of

different dynamical processes of the protein, and thus studying these processes in trehalose solutions can help elucidate the role of trehalose for protein stability. In order to investigate these processes over a broad temperature range, the water concentration was kept low (<35wt%) such that ice formation was inhibited, and the solvent entered the super-cooled regime, and subsequently into a vitrified state during cooling. The samples had either 7 or 11 waters per trehalose molecule, and contained various concentrations of myoglobin.

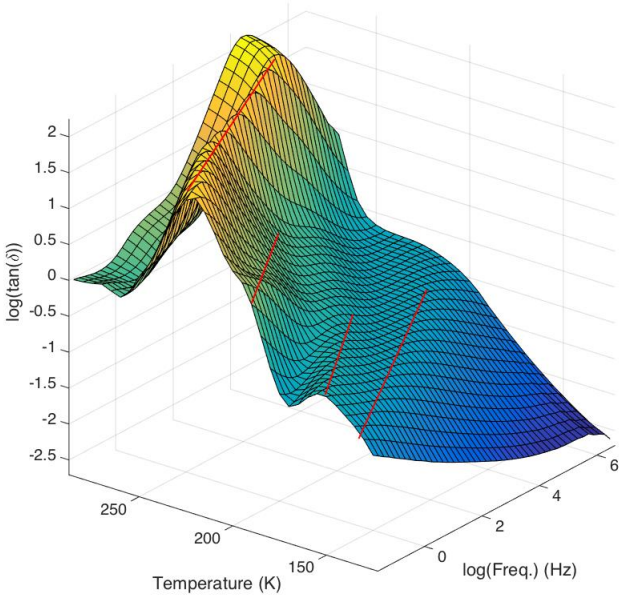


Figure 5.6: $\tan(\delta)$ for a sample of myoglobin in trehalose and water as a function of frequency and temperature. Red lines roughly indicate four different relaxation processes.

The results showed that there were three different local Arrhenius processes below T_g thought to be related to water, trehalose, and water coupled to local protein surface motions. In the system with the most trehalose, the water process was superimposed with the trehalose process. In the system with the most protein, the trehalose process was shown to disappear, probably due to macromolecular crowding effects. The protein-water process was shown to exhibit an Arrhenius to non-Arrhenius transition around the glass transition temperature, where it merges with the solvent α-relaxation. At the offset of the glass tran-

sition another, slower, non-Arrhenius process entered the frequency window, which is believed to originate from large-scale protein conformation changes. It is furthermore shown that this process exhibits a similar temperature dependence as the cooperative solvent relaxation, and it is therefore postulated that this slower protein process is slaved by the faster solvent relaxation. However, it should be noted that the fitting of the temperature range at and above T_g is not fully consistent with e.g. the calorimetric results. Therefore, the results related to the processes above T_g, such as the observed slaving behavior, are still preliminary and further investigation is needed.

6

Conclusions and Outlook

The particular conformation of a trehalose molecule causes the hydroxyl groups of the trehalose molecule to prefer interacting with water molecules (hydrogen bonding), rather than other trehalose hydroxyl groups (both internal and external). This means that trehalose avoids clustering and folding, yielding plenty of interactions with its environment (see e.g. Ref. 39 or paper I). However, this is not a unique property of trehalose; many different molecules provide a large amount of HB-sites. In paper II, for example, it was shown that sucrose exhibits quite similar hydrogen bonding to water, although trehalose perturbs the natural structure of water to a higher degree. However, it is unlikely that this property can explain the superior protective effects of trehalose. There ought to be some properties of trehalose beyond that, which have been extensively discussed in this thesis.

Trehalose tends to not bind directly to protein surfaces, but rather traps a layer of water at the protein surfaces instead (see e.g. Ref. 89 or papers III–V). This means that trehalose does not enforce its own structure on to the protein surface, which – if it occurred – could result in a destabilization of the entire protein. It was also shown that the presence of trehalose reduces the dynamics of the protein hydration water, and subsequently also reduce the dynamics of the protein, since certain protein motions, important for the function

of the proteins, have been shown to be driven by the dynamics of the hydration water.[91]

Another important property of trehalose is its ability to reduce protein-protein interactions (see e.g. Refs. 102, 103, and paper IV). This could be important for the role of trehalose as a stabilizing molecule, but possibly also important for the use of trehalose as a treatment against certain neurodegenerative diseases.[102,103] The reason for why trehalose is good at preventing protein aggregations is speculated to be due to that it surrounds the proteins homogeneously,[39,41] thus effectively shielding the protein molecules from one another via steric hindrance. However, the existence of such a trehalose layer, as proposed by e.g. Ref. 90, could not be identified from the results in this paper.

Future work on the role of trehalose for the purpose of protein stabilization will investigate the existence of such a trehalose layer, by the use of neutron diffraction and EPSR modeling. Furthermore, the results in papers III – VI only show how trehalose behaves in aqueous/hydrated protein solutions, and thus, the results by themselves do not necessarily determine if there is anything unique about trehalose. It would therefore be of interest to compare to e.g. sucrose in order to determine if there truly is any special aspect regarding the stabilizing role of trehalose.

A more complete picture of how trehalose acts in interaction with biological materials, at preservation conditions, may result in a more general understanding of biomolecular stabilization mechanisms. This can lead to improved cryopreservation and lyophilization procedures, which would be vastly beneficial to a large number of fields.

Acknowledgements

I would like to start by thanking my supervisor, Jan Swenson, for all his advice and guidance, and also for hiring me to work on this fascinating topic. I also want to thank my co-supervisor, Helén Jansson, for all her support and great discussions. Also, a big thank you to everyone who contributed to the writing and discussions with the scientific papers.

Thanks to all the people at KMF and Biological physics for making it fun to come to work. I would also like to thank Ezio for helping me with basically everything in the lab, and a special thanks to all the present and former members of the Swenson group: Khalid, Fredrik, and Alexandr for fun and interesting lunch-discussions and scientific support.

Finally, I want to express my gratitude to my friends and family for all your support, and of course especially to Anna for being an awesome companion and for adding well-needed silliness to my life.

Christoffer

Bibliography

[1] A. Liesowska, *Siberian scientists announce they now have a 'high chance' to clone the woolly mammoth*, `http://siberiantimes.com/science/casestudy/news/exclusive-siberian-scientists-announce-they-now-have\-a-high-chance-to-clone-the-extinct-woolly-mammoth/`, Accessed on: Jul 15, 2016.

[2] J. O. M. Karlsson and M. Toner, *Biomaterials*, 1996, **17**, 243–256.

[3] J. G. Day and G. Stacey, *Cryopreservation and freeze-drying protocols*, Springer Science & Business Media, 2007, vol. 368.

[4] D. E. Pegg, *Cryopreservation and freeze-drying protocols*, 2007, 39–57.

[5] G. Adams, *Cryopreservation and Freeze-Drying Protocols*, 2007, 15–38.

[6] P. Matejtschuk, in *Cryopreservation and Freeze-Drying Protocols*, ed. J. G. Day and G. N. Stacey, Humana Press, Totowa, NJ, 2007, pp. 59–72.

[7] A. H. Haines, *Organic & biomolecular chemistry*, 2006, **4**, 702–706.

[8] N. K. Jain and I. Roy, *Protein Science*, 2009, **18**, 24–36.

[9] S. R. Elliott, *Physics of amorphous materials*, Longman Scientific and Technical, UK, 1990.

[10] R. A. L. Jones, *Soft condensed matter*, Oxford University Press, 2002, vol. 6.

[11] A. K. Soper, *GudrunN and GudrunX - Programs for correcting raw neutron and x-ray total scattering data to differential cross section*, 2012.

[12] L. Van Hove, *Phys. Rev.*, 1954, **95**, 249–262.

[13] W. Kauzmann and D. Eisenberg, *The structure and properties of water*, Clarendon Press, 1969.

[14] A. H. Narten, M. D. Danford and H. Levy, *Discussions of the Faraday Society*, 1967, **43**, 97–107.

[15] H. E. Stanley, *MRS bulletin*, 1999, **24**, 22–30.

[16] A. K. Soper, *arXiv preprint arXiv:1411.1322*, 2014.

[17] A. K. Soper, *Chemical Physics*, 2000, **258**, 121–137.

[18] D. Dhabal, M. Singh, K. T. Wikfeldt and C. Chakravarty, *The Journal of Chemical Physics*, 2014, **141**, 174504.

[19] E. Duboué-Dijon and D. Laage, *The Journal of Physical Chemistry B*, 2015, **119**, 8406–8418.

[20] *ISIS Disordered Materials Database*, `http://www.isis.stfc.ac.uk/groups/disordered-materials/database/`, Accessed on: Jun 10, 2016.

[21] C. Colaço, S. Sen, M. Thangavelu, S. Pinder and B. Roser, *Bio/Technology*, 1992, **10**, 1007.

[22] M. Uritani, M. Takai and K. Yoshinaga, *The Journal of Biochemistry*, 1995, **117**, 774–779.

[23] T. Duong, R. Barrangou, W. M. Russell and T. R. Klaenhammer, *Applied and environmental microbiology*, 2006, **72**, 1218–1225.

[24] A. Wiemken, *Antonie van Leeuwenhoek*, 1990, **58**, 209–217.

[25] G. Bellavia, G. Cottone, S. Giuffrida, A. Cupane and L. Cordone, *The Journal of Physical Chemistry B*, 2009, **113**, 11543–11549.

[26] G. Bellavia, S. Giuffrida, G. Cottone, A. Cupane and L. Cordone, *Journal of Physical Chemistry B*, 2011, **115**, 6340–6346.

[27] D. Barreca, G. Laganà, S. Magazù, F. Migliardo, G. Gattuso and E. Bellocco, *International Journal of Biological Macromolecules*, 2014, **63**, 225–232.

[28] J. H. Crowe, L. M. Crowe, J. F. Carpenter and C. Aurell Wistrom, *Biochemical Journal*, 1987, **242**, 1–10.

[29] W. Q. Sun and A. C. Leopold, *Comparative Biochemistry and Physiology Part A: Physiology*, 1997, **117**, 327–333.

[30] L. M. Crowe, *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 2002, **131**, 505–513.

[31] N. Møbjerg, K. A. Halberg, A. Jørgensen, D. Persson, M. Bjørn, H. Ramløv and R. M. Kristensen, *Acta Physiologica*, 2011, **202**, 409–420.

[32] G. G. Birch, *Advances in carbohydrate chemistry*, 1963, **18**, 201–225.

[33] J. L. Green and C. A. Angell, *The Journal of Physical Chemistry*, 1989, **93**, 2880–2882.

[34] F. Sussich, C. Skopec, J. Brady and A. Cesàro, *Carbohydrate Research*, 2001, **334**, 165–176.

[35] D. Kilburn, S. Townrow, V. Meunier, R. Richardson, A. Alam and J. Ubbink, *Nat Mater*, 2006, **5**, 632–635.

[36] C. Olsson, H. Jansson and J. Swenson, *The Journal of Physical Chemistry B*, 2016, **120**, 4723–4731.

[37] A. M. Lammert, S. J. Schmidt and G. A. Day, *Food Chemistry*, 1998, **61**, 139–144.

[38] M.-O. Portmann and G. Birch, *Journal of the Science of Food and Agriculture*, 1995, **69**, 275–281.

[39] A. Lerbret, P. Bordat, F. Affouard, M. Descamps and F. Migliardo, *The Journal of Physical Chemistry B*, 2005, **109**, 11046–11057.

[40] S. Magazù, P. Migliardo, A. M. Musolino and M. T. Sciortino, *The Journal of Physical Chemistry B*, 1997, **101**, 2348–2351.

[41] M. Malferrari, A. Nalepa, G. Venturoli, F. Francia, W. Lubitz, K. Mobius and A. Savitsky, *Physical Chemistry Chemical Physics*, 2014, **16**, 9831–9848.

[42] L. Sapir and D. Harries, *The Journal of Physical Chemistry B*, 2011, **115**, 624–634.

[43] L. Sapir, D. E. Shalev and D. Harries, *Biophysical Journal*, 2011, **100**, 396a–397a.

[44] C. Branca, S. Magazù, G. Maisano and P. Migliardo, *The Journal of Chemical Physics*, 1999, **111**, 281–287.

[45] C. Branca, S. Magazù, G. Maisano and F. Migliardo, *Physical Review B*, 2001, **64**, 224204.

[46] C. Branca, S. Magazu, F. Migliardo and P. Migliardo, *Physica A: Statistical Mechanics and its Applications*, 2002, **304**, 314–318.

[47] C. Branca, V. Magazu, G. Maisano, F. Migliardo and A. K. Soper, *Applied Physics A*, 2002, **74**, s450–s451.

[48] C. Branca, S. Maccarrone, S. Magazù, G. Maisano, S. M. Bennington and J. Taylor, *The Journal of Chemical Physics*, 2005, **122**, 174513.

[49] S. Magazù, G. Maisano, F. Migliardo and C. Mondelli, *Biophysical Journal*, 2004, **86**, 3241–3249.

[50] S. Magazù, F. Migliardo and M. Telling, *Food Chemistry*, 2008, **106**, 1460–1466.

[51] S. E. Pagnotta, S. E. McLain, A. K. Soper, F. Bruni and M. A. Ricci, *The Journal of Physical Chemistry B*, 2010, **114**, 4904–4908.

[52] E. Shalaev and A. K. Soper, *The Journal of Physical Chemistry B*, 2016, **120**, 7289–7296.

[53] J. Teixeira, J.-M. Zanotti, M.-C. Bellissent-Funel and S.-H. Chen, *Physica B: Condensed Matter*, 1997, **234-236**, 370 – 374.

[54] J. Swenson and J. Teixeira, *The Journal of chemical physics*, 2010, **132**, 014508.

[55] A. K. Soper, M. A. Ricci, F. Bruni, N. H. Rhys and S. E. McLain, *The Journal of Physical Chemistry B*, 2018, **122**, 7365–7374.

[56] C. Mathews, K. Van Holde and K. Ahern, *Biochemistry. 3rd*, San Francisco, Calif.: Benjamin Cummings. xxviii, 2000.

[57] Image of 1MBN (Watson, H.C. (1969) The Stereochemistry of the Protein Myoglobin. *Prog.Stereochem.* 4: 299) created with Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/.

[58] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. Parrish, H. Wyckoff and D. C. Phillips, *Nature*, 1958, **181**, 662–666.

[59] F. Franks, *Water, a Comprehensive Treatise: The physics and physical chemistry of water*, Plenum Press, 1972.

[60] H. Frauenfelder, P. W. Fenimore, G. Chen and B. H. McMahon, *Proceedings of the National Academy of Sciences*, 2006, **103**, 15469–15472.

[61] P. W. Fenimore, H. Frauenfelder, B. H. McMahon and F. G. Parak, *Proceedings of the National Academy of Sciences*, 2002, **99**, 16047–16051.

[62] V. Lubchenko, P. G. Wolynes and H. Frauenfelder, *The Journal of Physical Chemistry B*, 2005, **109**, 7488–7499.

[63] J. A. Rupley and G. Careri, *Advances in protein chemistry*, 1991, **41**, 37–172.

[64] J. Swenson, H. Jansson, J. Hedström and R. Bergman, *Journal of Physics: Condensed Matter*, 2007, **19**, 205109.

[65] H. Jansson, R. Bergman and J. Swenson, *The Journal of Physical Chemistry B*, 2011, **115**, 4099–4109.

[66] C. A. Knight, A. L. De Vries and L. D. Oolman, *Nature*, 1984, **308**, 295–296.

[67] J. E. Lovelock, *Biochemical Journal*, 1954, **56**, 265.

[68] D. E. Pegg, Seminars in reproductive medicine, 2002, pp. 5–14.

[69] G. B. Strambini and E. Gabellieri, *Biophysical Journal*, 1996, **70**, 971–976.

[70] S. P. Leibo, J. J. McGrath and E. G. Cravalho, *Cryobiology*, 1978, **15**, 257–271.

[71] Z. W. Yu and P. J. Quinn, *Bioscience Reports*, 1994, **14**, 259–281.

[72] J. H. Crowe, J. F. Carpenter and L. M. Crowe, *Annual Review of Physiology*, 1998, **60**, 73–103.

[73] S. J. Prestrelski, T. Arakawa and J. F. Carpenter, *Archives of Biochemistry and Biophysics*, 1993, **303**, 465–473.

[74] S. J. Prestrelski, N. Tedeschi, T. Arakawa and J. F. Carpenter, *Biophysical Journal*, 1993, **65**, 661–671.

[75] J. Wolfe, *Functional Plant Biology*, 1987, **14**, 311–318.

[76] J. H. Crowe, J. F. Carpenter, L. M. Crowe and T. J. Anchordoguy, *Cryobiology*, 1990, **27**, 219–231.

[77] J. H. Crowe, S. B. Leslie and L. M. Crowe, *Cryobiology*, 1994, **31**, 355–366.

[78] I. Roy and M. N. Gupta, *Biotechnology and applied biochemistry*, 2004, **39**, 165–177.

[79] J. F. Carpenter and J. H. Crowe, *Biochemistry*, 1989, **28**, 3916–3922.

[80] N. Jovanović, A. Bouchard, G. W. Hofland, G.-J. Witkamp, D. J. A. Crommelin and W. Jiskoot, *European journal of pharmaceutical sciences*, 2006, **27**, 336–345.

[81] S. N. Timasheff, *Proceedings of the National Academy of Sciences*, 2002, **99**, 9721–9726.

[82] S. N. Timasheff, *Biochemistry*, 1992, **31**, 9857–9864.

[83] T. Arakawa and S. N. Timasheff, *Biochemistry*, 1982, **21**, 6536–6544.

[84] T. Arakawa and S. N. Timasheff, *Biophysical Journal*, 1985, **47**, 411–414.

[85] E. Y. Chi, S. Krishnan, B. S. Kendrick, B. S. Chang, J. F. Carpenter and T. W. Randolph, *Protein Science : A Publication of the Protein Society*, 2003, **12**, 903–913.

[86] L. M. Crowe, D. S. Reid and J. H. Crowe, *Biophysical Journal*, 1996, **71**, 2087–2093.

[87] K. Tanaka, T. Takeda and K. Miyajima, *Chemical & Pharmaceutical Bulletin*, 1991, **39**, 1091–1094.

[88] S. D. Allison, M. C. Manning, T. W. Randolph, K. Middleton, A. Davis and J. F. Carpenter, *Journal of Pharmaceutical Sciences*, 2000, **89**, 199–214.

[89] P. S. Belton and A. M. Gil, *Biopolymers*, 1994, **34**, 957–961.

[90] R. D. Lins, C. S. Pereira and P. H. Hünenberger, *Proteins: Structure, Function, and Bioinformatics*, 2004, **55**, 177–186.

[91] D. Corradini, E. G. Strekalova, H. E. Stanley and P. Gallo, *Scientific Reports*, 2013, **3**, 1218.

[92] S. Magazù, F. Migliardo and M. T. F. Telling, *European Biophysics Journal*, 2006, **36**, 163–171.

[93] P. Bordat, A. Lerbret, J. P. Demaret, F. Affouard and M. Descamps, *EPL (Europhysics Letters)*, 2004, **65**, 41.

[94] Y. Choi, K. W. Cho, K. Jeong and S. Jung, *Carbohydrate research*, 2006, **341**, 1020–1028.

[95] G. Cottone, *The Journal of Physical Chemistry B*, 2007, **111**, 3563–3569.

[96] S. Giuffrida, G. Cottone and L. Cordone, *Biophysical Journal*, 2006, **91**, 968–980.

[97] A. Cerami, *The role of the Maillard reaction in vivo*, The Royal Society of Chemistry: Cambridge, England, 1994.

[98] M. R. Brown, T. J. Keith and H. R. Knull, *Neurochemistry international*, 1992, **21**, 177–183.

[99] C. Schebor, L. Burin, M. d. P. Buera and J. Chirife, *LWT - Food Science and Technology*, 1999, **32**, 481–485.

[100] W. Wang, *International Journal of Pharmaceutics*, 2005, **289**, 1–30.

[101] C. A. Ross and M. A. Poirier, *Nature medicine*, 2004, **10**, S10.

[102] M. Tanaka, Y. Machida, S. Niu, T. Ikeda, N. R. Jana, H. Doi, M. Kurosawa, M. Nekooki and N. Nukina, *Nature medicine*, 2004, **10**, 148–154.

[103] R. Liu, H. Barkhordarian, S. Emadi, C. B. Park and M. R. Sierks, *Neurobiology of Disease*, 2005, **20**, 74–81.

[104] E. Y. Chi, S. Krishnan, T. W. Randolph and J. F. Carpenter, *Pharmaceutical Research*, 2003, **20**, 1325–1336.

[105] F. Béranger, C. Crozet, A. Goldsborough and S. Lehmann, *Biochemical and Biophysical Research Communications*, 2008, **374**, 44–48.

[106] R. K. Chaudhary, J. Kardani, K. Singh, R. Banerjee and I. Roy, *Neuro-Molecular Medicine*, 2014, **16**, 280–291.

[107] M. A. Singer and S. Lindquist, *Molecular Cell*, 1998, **1**, 639–648.

[108] S. James and J. J. McManus, *The Journal of Physical Chemistry B*, 2012, **116**, 10182–10188.

[109] A. Lerbret, F. Affouard, A. Hédoux, S. Krenzlin, J. Siepmann, M.-C. Bellissent-Funel and M. Descamps, *The Journal of Physical Chemistry B*, 2012, **116**, 11103–11116.

[110] J. J. Valente, K. S. Verma, M. C. Manning, W. William Wilson and C. S. Henry, *Biophysical Journal*, 2005, **89**, 4211–4218.

[111] M. Panzica, A. Emanuele and L. Cordone, *The Journal of Physical Chemistry B*, 2012, **116**, 11829–11836.

[112] S. W. Lovesey, *Theory of neutron scattering from condensed matter*, Oxford University Press, 1984.

[113] D. S. Sivia, *Elementary scattering theory: for X-ray and neutron users*, Oxford University Press, 2011.

[114] G. E. Bacon, *Neutron scattering in chemistry*, Butterworth-Heinemann Limited, 1977.

[115] T. Brückel, *Neutron Scattering: Lectures of the JCNS Laboratory Course Held at Forschungszentrum Jülich and the Research Reactor FRM II of TU Munich*, Forschungszentrum Jülich, 2010, vol. 27.

[116] A. K. Soper, *Molecular Physics*, 2009, **107**, 1667–1684.

[117] G. Placzek, *Physical review*, 1952, **86**, 377.

[118] J.-P. Hansen and I. R. McDonald, *Theory of simple liquids*, Elsevier, 1990.

[119] R. L. McGreevy and L. Pusztai, *Molecular Simulation*, 1988, **1**, 359–367.

[120] A. K. Soper, *Chemical Physics*, 1996, **202**, 295–306.

[121] A. K. Soper, *Empirical Potential Structure Refinement - EPSRshell - A Users Guide*, 2015.

[122] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845–854.

[123] J. Huang and A. D. MacKerell, *Journal of Computational Chemistry*, 2013, **34**, 2135–2145.

[124] W. Damm, A. Frontera, J. Tirado-Rives and W. L. Jorgensen, *Journal of Computational Chemistry*, 1997, **18**, 1955–1970.

[125] F. Kremer and A. Schönhals, *Broadband Dielectric Spectroscopy*, Springer Berlin Heidelberg, 2002.

[126] V. Raicu and Y. Feldman, *Dielectric Relaxation in Biological Systems: Physical Principles, Methods, and Applications*, Oxford University Press, 2015.

[127] S. Havriliak and S. Negami, *Polymer*, 1967, **8**, 161 – 210.

[128] D. T. Bowron, A. K. Soper, K. Jones, S. Ansell, S. Birch, J. Norris, L. Perrott, D. Riedel, N. J. Rhodes and S. R. Wakefield, *Review of Scientific Instruments*, 2010, **81**, 33905.

[129] S. E. Pagnotta, M. A. Ricci, F. Bruni, S. McLain and S. Magazu, *Chemical Physics*, 2008, **345**, 159–163.

[130] W. B. O'Dell, D. C. Baker and S. E. McLain, *PLoS ONE*, 2012, **7**, e45311.

[131] P. B. Conrad and J. J. de Pablo, *The Journal of Physical Chemistry A*, 1999, **103**, 4049–4055.

[132] W. Humphrey, A. Dalke and K. Schulten, *Journal of molecular graphics*, 1996, **14**, 33–38.

[133] C. J. Carlile and M. A. Adams, *Physica B: Condensed Matter*, 1992, **182**, 431–440.

[134] M. Adams, W. Howells and M. Telling, *The IRIS user guide*, Council for the central lab. of the research councils (clrc) technical report, 2001.

[135] M. Wübbenhorst and J. van Turnhout, *Journal of Non-Crystalline Solids*, 2002, **305**, 40 – 49.

[136] S. Cerveny and J. Swenson, *Phys. Chem. Chem. Phys.*, 2014, **16**, 22382–22390.

[137] V. Samouillan, D. Tintar and C. Lacabanne, *Chemical Physics*, 2011, **385**, 19 – 26.

[138] *WinFit Owner's manual.*, Novocontrol 2000.

[139] L. C. Thomas, *Making Accurate DSC and MDSC®Specific Heat Capacity Measurements with the Q1000 Tzero™DSC*, http://www.tainstruments.co.jp/application/pdf/Thermal_Library/Applications_Briefs/TA310.PDF.